

School Performance: Australian State/Territory comparisons of students' achievements in national and international studies

Ken Rowe

*Australian Council for Educational Research*¹

March 2006

Abstract: Policy activities related to outcomes and standards-based educational performance indicators and their links with growing demands for *accountability*, *standards monitoring*, *benchmarking*, *school effectiveness* and *reform* are widespread and well established in many countries throughout the world. While the long-term goals of school education may be expressed as the enhancement of young peoples' access to and participation in society, as well as preparation for meeting the constantly changing demands of the modern workplace, the most direct and readily accessible measures of student and school performance are obtained from assessments of students' academic achievements. Despite several limitations, achievement data obtained from both national and international studies have several benefits that include: (a) the potential to provide valuable information concerning student and school performance compared with other national systems, and (b) generate understandings (as well as raise questions) about observed differences among educational jurisdictions – within and between countries. To this end, the present paper presents findings from analyses of students' achievements in *Literacy*, *Mathematics* (or *Numeracy*) and *Science* – obtained from participation in national and international studies, and compares students' achievements at the student and school levels, and between Australia's eight States and Territories. Implications of the findings are discussed.

Introductory comments

The provision of schooling is one of the most massive and ubiquitous undertakings of the modern state. Schools account for substantial proportions of public and private expenditure, and are universally regarded as vital instruments of social and economic policy aimed at promoting individual fulfilment, social progress and national prosperity. It has long been recognised that the key to such prosperity at both the individual and national level is the provision of quality schooling. Since schooling generates a substantial quantity of paid employment for teachers and administrators, it is not surprising that there has long been an interest in knowing how *effective* the provision of school education is and how it can be improved.²

The global economic, technological and social changes under way, requiring responses from an increasingly skilled workforce, make high quality schooling an imperative. Whereas OECD education ministers have recently committed their countries to the goal of raising the quality of learning for all, this ambitious goal will not be achieved unless all children, irrespective of their characteristics, backgrounds and locations, receive high-quality schooling and teaching in particular (OECD, 2005a,b).

¹ Correspondence related to this paper should be directed to: Dr Ken Rowe, Research Director, Learning Processes and Contexts Research Program, ACER, 347 Camberwell Road (Private Bag 55), Camberwell, VIC 3142, Australia; Tel: +61 3 9835 7489; Email: rowek@acer.edu.au.

² See, for example: Bosker, Creemers and Scheerens (1994); Caldwell (2003); Coleman *et al.* (1966); DES, (1984); Edmonds (1979a,b, 1981); Goldstein (1997, 2001); Goodlad, (1982, 1983); Hanusheck (2004, 2005a,b,c); Hanusheck *et al.* (2005); Hanusheck and Jorgenson (1996); Hanusheck and Raymond (2004); Hanushek, Rivkin and Kain (2005); Jencks *et al.* (1972); Masters (1990, 1991, 1994); Mortimore (1991, 1992, 1995, 2001); NCEE, 1997; OECD (1983, 1986, 1989, 2001); Podgursky *et al.* (2004); Reynolds, Hopkins and Stoll (1993); Rowe (2001a,b, 2003, 2004a,b, 2005a,b,c); Rowe and Lievesley (2002); Rowe, Turner and Lane (2002); Rutter *et al.*, (1979); Scheerens and Bosker (1997); Willms (2000); Wyatt (1994).

Central to the goal of providing quality schooling has been the development, construction and increasing use of educational *performance indicators*. The work of Professor Eric Hanusheck and colleagues at Stanford University (USA) continues to make noteworthy contributions to understandings about *economic indicators of quality schooling* (for several references to this on-going work, see relevant citations given in footnote 2). A brief outline of what is entailed in such indicators is helpful.

Performance indicators

In general, *performance indicators* (PIs) are defined as *data indices of information* by which the functional *quality* of institutions or systems may be measured and evaluated. Typically, within the context of specified goals and objectives, PI data are ‘measures’ of operational and functional aspects of organizations and/or systems, and provide evidential bases for determining the extent to which such goals and objectives have and are being met. PIs serve various purposes, the most notable of which are for **monitoring, feedback, policy formulation, target-setting, evaluating** and **reforming**. Although the essential features of educational PIs are consistent with their counterparts in other government and corporate enterprises, they also have unique characteristics – key aspects of which have been highlighted by Rowe (2001a, 2004a, 2005a) and by Rowe and Lievesley (2002). At the outset, however, it is helpful to note the importance of educational PIs in prevailing local and international contexts.

The nature and purpose of educational PIs

During the last thirty years, education systems throughout the world have been subject to considerable reform and change – all justified on the grounds (or at least the rhetoric) of **improving** the *quality* of school education. A key feature of this change has been the frequent revisions of style and policy focus, especially in respect of PIs, with major emphases being placed on the assessment and monitoring of student learning outcomes – mostly in Literacy, Numeracy (Mathematics) and Science. Indeed, current policy activities related to outcomes and standards-based educational PIs and their links with growing demands for *accountability, standards monitoring, benchmarking, school effectiveness* and *reform* are widespread and well established in many developed countries (e.g., Buckingham, 2003; Chapman *et al.*, 1991; Dorn, 1998; Hill & Cr  vola, 1999; Forster, Masters & Rowe, 2001; Rowe, 2001a, 2005a; Tucker & Coddling, 1998; Visscher & Coe, 2002; Willms, 2000).

Such emphases are aptly illustrated in the reported proceedings of a meeting under the auspices of the *Summit of the Americas* (2002), which states:

Although it is now part of daily life in schools and in debates between specialists, education assessment has recently become a relevant topic for governments and society, especially because of the economic crisis and the acceleration of the globalization process, which made investments in education a strategic point while the resources available for the sector have shrunk.

In many developed countries including Australia, much of this activity has been (and continues to be) focussed on linking *inputs* and *processes* of educational systems (e.g., physical resources and curriculum provision) with *outputs* (e.g., improvements in student achievement outcomes, as well as in school and system performance). A major effect of such activity has been to signal government policy intention to:

- encourage system accountability to ensure both efficient and effective utilization of resources, and
- bring the delivery of educational services into public sector accounting, underscored by a concern to ensure that such services represent ‘value for money’.

Whereas the provision of quality education is critical to the development of all countries, it is especially the case for developing countries where there is considerable pressure to increase access to education, but not at the expense of quality. Hence, the demand is to ensure that PIs

(including assessments of students' learning and achievements) do not provide a partial, and thus potentially misleading picture of either *quality* or *effectiveness*, as has often been the case.

Despite the difficulties entailed in defining *educational effectiveness* at the school, system, national and international levels, and reaching consensus on the relevant criteria, a good deal of discussion has focused on what is meant by *quality schooling* and *quality teaching*, and how they might be measured and improved. Although the term *quality* is likewise problematic, the "...measurement of the *quality* of schooling is of critical importance at a time when so much school reform in so many parts of the world is being undertaken" (Mortimore, 1991, p. 214). In fact, concerns about the *quality* of school education and its monitoring have long been high priority policy issues in all OECD countries (OECD, 1983, 1986, 1989, 1993, 1995, 2001, 2005a,b). An illustration of this priority is evident in the assertion by Manno (1994):

When judging educational quality, either we focus on what schools spend – or one of its many variants – or we focus on what students achieve, what they know and can do. Those who advocate a focus on outcomes in judging educational quality hold one common belief: we must specify what we expect all children to learn, and we must assess them to determine whether they have learned it.

While the long-term goals of school education may be expressed as the enhancement of young peoples' access to and participation in society, as well as preparation for meeting the constantly changing demands of the modern workplace (OECD, 1983, 1986, 2005a), the most direct and readily accessible measures of schooling outcomes are obtained from assessments of students' academic attainments. Herein, however, lies a dilemma that is evidenced in strident critiques of traditional and prevailing psychometric models for test and examination modes of assessment (e.g., Berlak, 1992) and an equally strident chorus of concern for the deleterious effects of test-driven and 'test-dominated' curricula (e.g., Kellaghan, Madaus & Airasian, 1992; Lacey & Lawton, 1981). As Watson (1996) noted: "In high stakes testing environments, educational practitioners are likely to distort their behaviour in order to meet the demands of the indicator, usually to the detriment of their real job" (p. 119). Nisbet (1993, p. 25) further highlighted this dilemma in the following terms:

In today's schools, assessment is a main influence on how pupils learn and how teachers teach. Whether assessment is in the form of examinations and tests, or marks and grades for coursework, its influence is pervasive. Often it distorts the process of learning through teaching to the test, cramming, short-term memorising, anxiety and stress – to the extent that learning to cope with assessment has become almost as important as the genuine learning which such assessments are supposed to measure. For many young people, assessment dominates education.

Although measures of student learning and achievement outcomes are prime PIs of education systems and the services they provide and for which they are responsible, there are many others (including both *inputs* and *processes*) that constitute useful bases for informed planning and decision-making, followed by implementation and reform. If decisions for improvement are to be **data-informed**, rather than based on whim or ideology, then useful, dependable and timely information on PIs is required. Indeed, such bases constitute key purposes of specifying, gathering and using PIs for educational change and reform. In particular, PI information allows systems and their constituent organizational elements to: (1) formulate strategic policy priorities and their related targets, (2) specify achievable objectives, (3) implement them, and (4) evaluate the extent to which those target objectives have been attained.

The benefits and limitations of national and international assessment programs for monitoring student learning and achievement outcomes

The benefits and limitations of national and international monitoring programs for student learning and achievement outcomes have been well documented and require little reiteration here (e.g., Beaton *et al.*, 1999; Forster, 2000, 2001a,b; Goldstein, 2001, 2004; Greaney & Kellaghan, 1996; McGaw, 1991; Murphy *et al.*, 1996; Plomp, 1999; Rowe, 2004a; Rowe & Lievesley, 2002; Scheerens & Bosker, 1997; Visscher *et al.*, 2000; Willms, 2000). In brief, the

benefits of national assessments include the provision of systematic and regular measures of student learning and achievement outcomes. They are designed to evaluate the relative 'health' of education systems, to monitor achievement across the systems, and provide information that allow comparisons of performance within the system of sub groups of students, as well as within and between districts, regions and states. The data obtained assist policy makers to allocate resources designed to maximize learning opportunities and outcomes for all students. Nonetheless, McGaw (1991, p. 138) pointed out both the benefits and risks involved in national achievement monitoring programs in the following terms:

The benefit of assessing all students is that each school obtains information about its program and teachers obtain potentially helpful diagnostic information about all students. The risk is that the universality of such a program will allow and even encourage comparisons among schools, without consideration of the effect of non-school factors on scores, and so oblige schools to concentrate more upon specific preparation for the tests.

Subsequently, and consistent with the warnings of Goldstein and Spiegelhalter (1996) about the dangers of publishing student and school performance data in the form of 'league tables', Rowe (2000, p. 92) observed:

The existence of an accountability climate that insists on providing published information that invites comparative judgements about the relative 'worth' of schools – and, inevitably, about the teachers who work in them – is problematic. It is a social and political minefield that has the potential for considerable harm unless it is handled with **great care**. Again, this is not to deny the usefulness of school-level educational performance indicators involving student achievement data, provided that relevant contextual factors have been taken into account and that the statistical uncertainty associated with the estimates obtained are displayed prominently.

While there are distinct advantages in implementing assessment programs at the beginning of the school year for *diagnostic* purposes to assist teachers in meeting the specific learning needs of students (both at the individual and cohort levels), as in France (see OECD, 1993), accountability pressures on State and Federal governments in Australia to monitor educational standards are political realities, and ones that are not likely to diminish. In this context, Hill (1995, p. 4) noted:

...accountability pressures have forced most education systems to press ahead with large-scale assessment programs. All government school education systems in Australia ... now operate programs to monitor educational standards. ... The principal motivation behind current assessment programs is to meet public demands for educational systems to be accountable for maintaining and indeed improving standards. As such, they tend to command broad support from the community, but rarely receive enthusiastic support from the teaching profession.

Achievement data obtained from both national and international studies have several benefits. In the case of international studies, given that the measurements of students' achievements are calibrated on common scales, such benefits include: (a) the potential to provide valuable PI information about a country's education system(s) in relation to other national systems concerning the performance of students and schools, and (b) generate understandings (as well as raise questions) about observed differences in the achievements of students from different educational systems. For example, Plomp (1999, pp. 1-2) has noted:

The understandings we obtain from cross-national comparisons of such policies as age of school entry, hours and methods of instruction, and teacher training, can provide us with new insights into the performance of our own educational system in general, and of the relationship between student performance and its antecedents and consequences in particular.

Findings from international studies of student achievement also have the advantage of attracting political and media attention. Thus, poor results can provide policy makers with a strategic rationale for intervention and budgetary support advocacy throughout education systems and their constituent jurisdictions (see: Forster, 2000, 2001a,b; Greaney & Kellaghan, 1996; Rowe & Lievesley, 2002). However, several studies have now shown that there are serious and *inherent* limitations to the usefulness of such indicators for providing reliable

judgements about educational institutions (e.g., Goldstein & Thomas, 1996; Goldstein & Spiegelhalter, 1996; Marsh, Rowe & Martin, 2002; Rowe, 2000, 2004a; Visscher *et al.*, 2000). Key reasons for these limitations are as follows:

- Against the background of what is known about *differential school effectiveness* (e.g., Nuttall, Goldstein, Prosser & Rasbash, 1989) it is not possible to provide simple summaries that capture all the important features of schools (see also: Bosker, Creemers & Scheerens, 1994; Hill & Rowe, 1996, 1998; Rowe, 2000, 2004a; Rowe & Hill, 1998; Rowe & Rowe, 1999; Rowe, Turner & Lane, 2002; Rowe & Stephanou, 2001, 2003; Scheerens & Bosker, 1997; Stephanou & Rowe, 2002; Visscher *et al.*, 2000; Willms, 2000).
- By the time information from a particular school has been analysed, it refers to a ‘cohort’ of students who entered that school several years previously so that its usefulness for *future* students and the making of judgements about *school effectiveness* may well be dubious. Moreover, where information is analysed on a yearly basis, it is necessary to make adjustments for prior contributing factors that extend over two or more years. In fact, it is increasingly recognised that schools, or teachers within those schools, should not be judged by a single ‘cohort’ of students, but rather on their performance over time (e.g., Goldstein, 1997; Thomas *et al.*, 1997; Thomson *et al.*, 2005). As noted by Goldstein (1997), this makes the historical nature of *school effectiveness* judgements an acute problem.
- Above all, even when suitable adjustments for students’ intake characteristics and prior achievements have been taken into account, even the resulting *value-added* estimates have too much *uncertainty* attached to them to provide reliable rankings. This point, illustrated elsewhere, is vital and one that is all too-frequently ignored by advocates of published ‘league tables’ (see: Rowe, 2000, 2004a; Rowe & Stephanou, 2001, 2003; Rowe, Turner & Lane, 2002).

Limitations of findings summarised in this paper

The limitations of findings from analyses of data derived from national and international assessment programs for monitoring student learning and achievement outcomes, as outlined above, apply equally to those presented in this paper. Principal among these is the limited number of available explanatory variables at the student-level and group-membership levels (e.g., class, school and State/Territory) to provide effect estimates required for adjustment. For example, the fact that Australian students (and their parents) are not obliged to disclose their ethnic (and religious) affiliations in surveys of any kind results in large proportions of ‘missing data’ for these variables. This is particularly relevant to obtaining effect estimates for students’ Indigenous status (i.e., Aboriginal and Torres Strait Island ‘membership’, or otherwise). Indeed, there is always the difficulty that any statistical model used to provide effect estimates at the student, contextual and group-membership levels, will fail to incorporate **all** the appropriate adjustments, or in some other way may be mis-specified. Thus, at best, effect estimates can only be used as ‘screening devices’ to identify ‘outliers’ (which could form the basis for follow-up), but they cannot and should not be used as definitive measures of the effect of those schools (or jurisdictions) on student learning *per se*.

A further limitation that applies to the present paper is the restriction on providing cross-sectoral comparisons of student achievement outcomes (i.e., across government, Catholic and independent schools). This restriction derives from directives by national Steering Committees for non-disclosure of cross-sectoral comparisons of findings from both national monitoring programs and international studies. Regardless of justifications for these directives, it is important to note that the results presented here lack adjustments for this major source of contextual variation – particularly given the increasing student enrolment ‘drift’ from government to non-government schools during the last decade in all Australian States and

Territories since 2000.³ Further, this deficiency imposes restrictions on estimating the differential effects of teaching and learning provision within and between sectors (see: Cuttance, 2001; Darling-Hammond, 2000; Hanushek, Rivkin & Kain, 2005; Hill & Rowe, 1996; Muijs & Reynolds, 2001; Rowe, 2004b).

Focus of the present paper

This paper focuses on the relative achievements of students located in schools throughout Australia's six States and two Territories. Hence, following in this section, comparative findings from analyses of the available data are presented – mostly for students' achievements in *Literacy, Numeracy* and *Science*. These data have been obtained from:

- system-wide, full cohort, monitoring programs of Year 3, 5 and 7 student 'benchmark' achievements in *Reading, Writing* and *Numeracy* from 1999 to 2004⁴ (Section 1);
- Australia's participation the OECD *Programme of International Student Assessment* (PISA) with a focus on performance in *Reading Literacy, Mathematical Literacy* and *Scientific Literacy* during 2000 and 2003, for stratified, representative samples of 15-year-old students – typically in their eleventh year of schooling – drawn from government, Catholic and independent schools (Section 2); and
- Australia's participation in the IEA *Third International Mathematics and Science Study* (TIMSS) of performance in *Mathematics* and *Science* during 2003, among a stratified, representative sample of Grade 8 students – typically in their ninth year of formal schooling – drawn from government, Catholic and independent schools (Section 3).⁵

For ease of reporting in the body of this paper, the technical and graphical findings are presented in **Appendices**.⁶ Hence, only summary comments derived from the related analyses and statistical modeling of the available data are provided here.

1.0 State/Territory comparisons of student achievement: Benchmark data for Literacy and Numeracy

The graphical summary of comparative findings that are presented in **Appendix A1** are for eight Australian State/Territories of Grade 3, 5 and 7 student 'benchmark' achievements from participation in system-wide Literacy and Numeracy monitoring programs. To assist interpretation, the key findings of available data between 1999-2003 are presented in graphical form – drawn from tabulated summaries reported by MCEETYA (2005).

Each graph, across the Australian States/Territories and by calendar year, provides percentage mean-point estimates of students achieving nationally agreed 'benchmark' performance standards, despite reservations of 'strict' comparability based on *post hoc* equating. Above all, it is important note that these 'benchmarks' are **minimum** standards – below which students' performances are deemed to be 'unacceptable' for their age/grade level of schooling. The major rationale for reporting such 'benchmarks' is to alert schools and systems to the need for interventions at the individual and group cohort levels.

³ See ABS (2006) for evidence of this 'drift' during the period 1980-2005.

⁴ To date, of the eight Australian State and Territory education systems, the ACT and TAS are the only government systems that annually assess full cohorts of Grade 9 students' literacy and numeracy achievements. Thus, for comparative purposes, only the 'benchmark' profiles for Grades 3, 5 and 7 are presented here (see Appendix A1).

⁵ TIMSS also assesses the *Mathematics* and *Science* performances of Grade 4 students (typically in their fifth year of schooling). However, in the interest of parsimony, this paper reports the relevant achievement data for Grade 8 students only.

⁶ Due to their efficiency and high-resolution graphics capabilities, the data analyses and graphical presentations of the findings were undertaken using *STATISTICA* (StatSoft, 2005) and *MLwiN* (Browne, Healy, Cameron & Charlton, 2005).

1.1 Summary of findings from Benchmark data

Despite minor variations across the specified calendar years given in Appendix A1, the *Literacy* and *Numeracy* ‘benchmark’ achievement profiles for Year 3, 5, and 7 students enrolled in ACT, NSW, VIC, QLD, SA, WA, and TAS schools were not significantly different. The performances of students enrolled in NT schools were mostly significantly below those of their counterparts in the other States and Territories. Two differences, however, are noteworthy: (a) during 2002, the percentage of Grade 3 students in NT schools who were at or above the national ‘benchmark’ for *Numeracy*, were not significantly different from their counterparts in WA (Fig. A1.3); and (b) between 2001 and 2003 a significantly **lower** percentage of Grade 5 students in both QLD and NT performed at or above ‘benchmark’ standards for *Reading* than their counterparts in the other States/Territories (Fig. A1.4).

It is important to note that PI information of the kind presented in Appendix A1 has limited utility since interpretations from raw, unadjusted percentage data of these kind are misleading (see: Rowe, 2001b, 2004a; Rowe & Stephanou, 2001, 2003; Stephanou & Rowe, 2002). That is, because such data are unadjusted for students’ ‘intake’ characteristics and contextual influences such as age, gender, home background factors such as family socioeconomic status (SES), etc., as well as student compositional characteristics at the school and system levels, responsible State/Territory comparisons of relative performance based on such data are not possible.⁷

Above all, results such as these give rise to queries about the adequacy of the post-hoc equating procedures implemented to ‘ensure’ that the constituent assessment items from year-to-year (and their transformed scale scores) have been calibrated on common scales for *Reading*, *Writing* and *Numeracy*. These procedures cast doubt on the *measurement adequacy* of the ‘benchmark’ scales to the extent that comparative judgements of student achievement within and between States/Territories from year-to-year are not viable.

Similarly, in the case of attempts to provide ‘value-added’ estimates, evaluations of student achievement ‘growth’ (albeit cross-sectional) are not possible. Under such circumstances, any adjustments that might be made to account for students’ background, intake’ and school compositional characteristics would be misleading. In sum, these findings underscore the importance and urgency for a national approach to the monitoring of student achievement outcomes. Related historical data are informative.

Evidence from the 1996 *National School English Literacy Survey* (Masters & Forster, 1997a,b) indicated that the proportion of Grade 3 and Grade 5 students in Australian schools who did not meet minimum performance standards of reading required for effective participation in further schooling was estimated to be as high as 27 per cent at Grade 3, and 29 per cent at Grade 5 (Masters & Forster, 1997b, p. 15).⁸ In 2003, the percentages of Australian students not achieving the minimum *National Benchmarks* for *Reading* were: ~8 per cent (Grade 3) and ~11 per cent (Grades 5 and 7).⁹ Despite apparent improvements since 1996, these outcomes are unacceptable in terms of the educational, psychosocial wellbeing and life chances of these Australians, as well as the economic and social future of the nation.

⁷ Rather than adjusting for students’ ‘intake’ and compositional characteristics at the school and State/Territory levels, the MCEETYA (2005) document merely provides separate percentage ‘benchmark’ estimates for male, female and ATSI students for each of the eight States and Territories.

⁸ Comparative international data are of interest. From the evidence cited in the report by British House of Commons Education and Skills Committee (2005), *Teaching Children to Read*, it is estimated that approximately 20 per cent of 11-year-old children in British schools do not achieve expected success in reading for their age. According to the National Center for Educational Statistics (US), 38 per cent of fourth graders (~9 year-olds) cannot read at a basic level – that is, they cannot read and understand a short paragraph similar to that in a children’s story book (Lyon, 2003, p. 1).

⁹ See MCEETYA (2005) and the graphical presentation of the data summarised in Appendix A1 of this paper.

2.0 State/Territory Comparisons of Students' Achievements in PISA

The comparative findings presented in **Appendix A2** across eight Australian State/Territories of student achievement outcomes derive from participation in two data-collection phases of the OECD *Programme of International Student Achievement* (PISA) to date, for 15-year-olds during 2000 and 2003 (see OECD, 2002, 2003). Due to variations across Australia's States and Territories in respect of school starting ages (and hence, age/grade membership), adjustments are made for participating students' *Age* (in years and months) and *Grade* (Grade level, or the number of years of formal schooling). Whereas the largest source of variation in school performance is typically attributed to differences in what students bring to school: their abilities and attitudes, and family and community wealth and background, the research evidence shows that school systems differ in the extent to which students' 'intake' characteristics and socio-economic (SES) background influences achievement (Marks, 2005; OECD, 2002).

The related research findings also show that there is often substantial variation in student performance within- and between-schools serving similar socio-economic catchment areas, as well as between classes within the same school (e.g., Hill & Rowe, 1996, 1998; Marks, 2000, 2005, 2006; Rowe, 2001b, 2003, 2004a). These differences imply that school system policies, and individual school and teacher practices, do make a difference in influencing student learning and achievement outcomes. Thus, adjustments are also made for student *Gender*, *SES* and *Home Educational Resources* (HEDRES). For specific details and definitions of both achievement and presage variables relevant to the PISA studies, see: Lokan, Greenwood and Cresswell (2001); Thomson, Cresswell and De Bortoli (2004).

To assist interpretation, findings from fitting single-level models to the student achievement data are compared with those from fitting multilevel models to the relevant data. Regrettably, too often within Australia (as well as in many other countries), the clustering effects of students-within schools are ignored by data analysts. In consequence, the resulting aggregation bias yields misestimates of the effects at best, and misleading findings at worst (Rowe, 2004a). In contrast, findings from multilevel analyses of the data are more 'responsible' since they more accurately reflect the ecological reality of students being nested within-classes and schools, such that major sources of variation may be identified and estimated. The estimation of these effects are not possible from fitting single-level models.

2.1 Summary of key findings from PISA 2000 and PISA 2003

Prior to commenting on the comparative findings from fitting single level and multilevel models to the PISA 2000 and PISA 2003 data presented in **Appendix A2**, it is useful to examine the functional relationship between students' achievement scores for *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy*. For this purpose, these relationships are provided using the PISA 2003 score data. Table 2.1 below presents the relevant correlations (by student Gender) for the total Australian sample, and Figure 2.1 provides the linear surface plots for males and females that illustrate the functional linear relationships.

Table 2.1 Relationships Between PISA 2003 Scaled Scores for Males and Females

Sub-Sample	Relationship	Correlation, r	Coefficient of Determination, r^2 (% of variance)
Males (n = 6335)	Reading & Maths	0.817	0.677 (67.7%)
	Reading & Science	0.880	0.774 (77.4%)
	Maths & Science	0.862	0.743 (74.3%)
Females (n = 6216)	Reading & Maths	0.796	0.634 (63.4%)
	Reading & Science	0.868	0.753 (75.3%)
	Maths & Science	0.842	0.709 (70.9%)

Note: all correlations are statistically significant beyond the $p < 0.0001$ level.

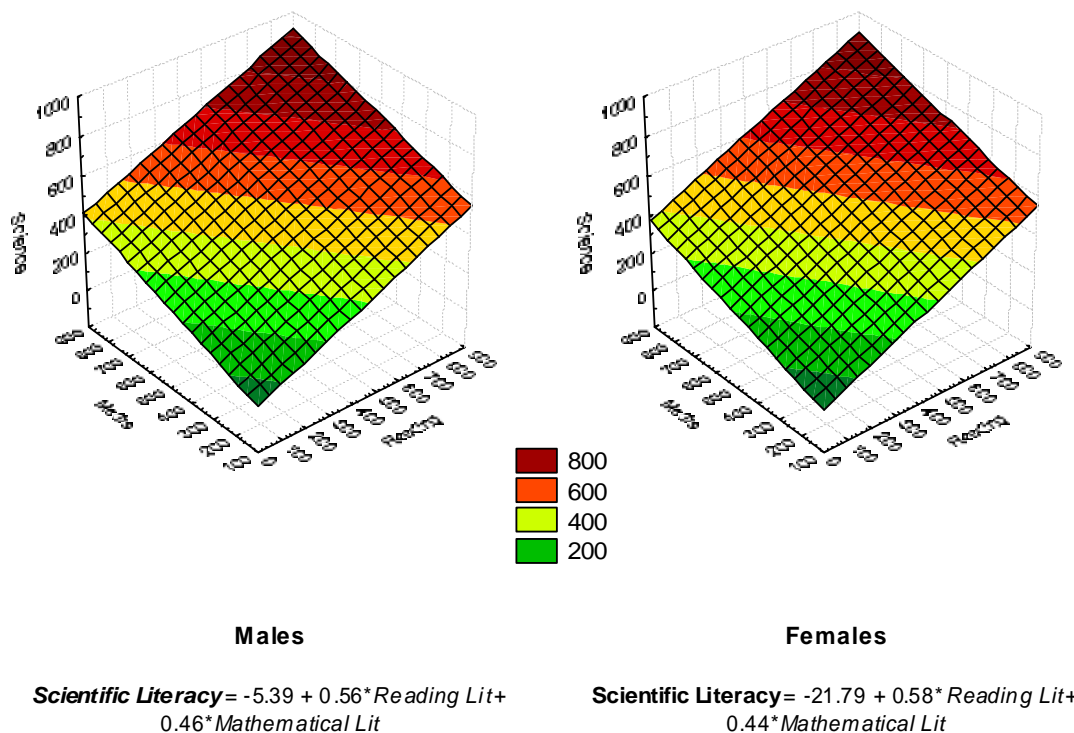


Figure 2.1 2D surface plots illustrating relationships between PISA 2003 scaled scores for *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy*, by student gender, with prediction equations for *Scientific Literacy*

Comment: The data summarised in Table 2.1 and illustrated in Figure 2.1 indicate strong functional relationships between students' achievement scores for PISA 2003 *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy*. Of particular interest was the importance of *Reading Literacy*¹⁰ as a predictor of students' scores for *Scientific Literacy*. For example, from Figure 4.2 in case of female students this yielded a 0.58 SD effect size, compared with 0.44 SDs for *Mathematical Literacy*. In the case of male students' *Scientific Literacy* scores, *Reading Literacy* yielded a similarly larger predictive effect size (0.56 SDs) than did *Mathematical Literacy* (0.46 SDs).

Notwithstanding the high level of *verbal processing* demand required by such assessments, these findings emphasise the importance of *Reading Literacy* in most areas of the school curriculum and its assessment – beginning with the pre-school and early years of formal schooling (see Rowe, 2005c). That is, *Reading Literacy* competence constitutes the foundational skill that underlies effective engagement with the school curriculum. Furthermore, the findings are supported by the work of Nobel Prize winning economist James Heckman's (2000, 2005) overview of the economic aspects of human skills formation. Heckman concludes that investment in the learning development of children and young people is crucial. For Heckman, literacy competence is an essential area of learning investment in the young, being a

¹⁰ In the OECD *Programme for International Student Assessment* (PISA), the concept of *Reading Literacy* emphasises skill in using written information in situations that students may encounter in their life both at and beyond school. Thus, *reading literacy* is defined as: '... understanding, using and reflecting on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society' (OECD, 2003, p. 108). For specific details related to the PISA 2000 and 2003 results relevant to Australia, see: Lokan, Greenwood and Cresswell (2001); Thomson, Cresswell and De Bortoli (2004).

‘skill that begets many other skills’ (an index of ‘self-productivity’, as he calls it), because it constitutes a ‘key part of our capacity to increase our capacity’.

From the comparative findings of fitting single level and multilevel models to the PISA 2000 and PISA 2003 data presented in **Appendix A2**, four features are worthy of emphasis. First, on average, the achievement performances of Australian 15-year-old students in *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy* have consistently been significantly above the OECD averages. As a nation, this result is to be celebrated – reflecting positively on the quality of teaching and learning provision in Australian schools – particularly for *Reading Literacy*. Of concern, however, is the increased ‘gap’ between 2000 and 2003 that separated the mean *Reading Literacy* achievements of male and female students (in favour of females). The plots of adjusted means given in Figure A2.1 (PISA 2000) compared with those given in Figure A2.9 (PISA 2003) provide evidence of this increased ‘gap’ between males and females in all Australian States and Territories.

Second, although these international assessments of *Reading Literacy* during 2000 and 2003 indicated that 15-year-old students in Australian schools performed notably better (on average) than the majority of their counterparts in other OECD countries, there have been notable variations between States/Territories and sub-groups of students. For example, 12 per cent of students (ACT, WA) to 28 per cent (NT) had not developed the literacy skills needed for further education, training and work (defined as *low achievers*), particularly indigenous students (35%) and males (17%). Similar proportional estimates have been reported for achievement in reading comprehension of 14-year-old Australian students between 1975 and 1998, and, with few exceptions, the estimates have remained constant during the period.¹¹ Furthermore, approximately 20 per cent of Australians aged 15-74 years have been identified as having “very poor” literacy skills, with an additional 28 per cent who “could be expected to experience some difficulties in using many of the printed materials that may be encountered in daily life” (ABS, 1997, p. 7). The importance of competence in reading for achievement in science and mathematics has already been noted and illustrated above (see Fig. 2.1) – low performance in which severely limits opportunities for further education and training, as well as active participation in economic and social life.

Third, the comparative State/Territory findings from fitting single-level explanatory models to the PISA 2000 and PISA 2003 student achievement data presented in Appendix A2, suggest significant variation between the States and Territories, following adjustments for students’ background and ‘intake’ characteristics. At the *prima facie* level, these findings indicate that the performances of students in ACT schools compared favourably with their counterparts in the other States and Territories. Indeed, in several instances, the achievement performances of students in ACT schools for *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy* were notably ‘better’ than their counterparts in other States and Territories. Nevertheless, the results from fitting more ‘responsible’ multilevel models to the achievement data (with similar adjustments), clearly indicated that these differences were inflated and hence, misleading.

A fourth finding worthy of highlight is the increased variation in the adjusted mean student achievement performances among schools in all States and Territories between 2000 and 2003. Findings from fitting single and multilevel models to the student achievement data for both PISA *Reading Literacy* and *Mathematical Literacy* indicated increased between-school residual variation from 2000 to 2003. For example, in the case of fitting single-level models to the relevant data for *Reading Literacy*, compare the plots given in Figures A2.1 (PISA 2000) and A2.9 (PISA 2003). Similarly, compare the residual plots from fitting multilevel models to the relevant data presented in Figures A2.5 (PISA 2000) and A2.14 (PISA 2003).

¹¹ See Rothman (2002), who notes: “For some groups, there has been improvement, most notably for students from language backgrounds other than English. For other groups, however, results indicate a significant achievement gap. The most significant gap is between Indigenous Australian students and all other students in Australian schools” (p. ix).

3.0 State/Territory Comparisons of Students' Achievements in TIMSS

The comparative findings presented in **Appendix A3** across eight Australian State/Territories of student performance in *Mathematics* and *Science* derive from participation in the IEA *Third International Mathematics and Science Study* (TIMSS) for students in their ninth year of formal schooling during 2003 (see: Martin *et al.*, 2004; Mullis *et al.*, 2004). As for analyses of the PISA data, due to variations across Australia's States and Territories in respect of school starting ages (and hence, age/grade membership), adjustments of the TIMSS data are made for participating students' *Age* (in years and months). Adjustments are also made for student *Gender*, *SES* and *Home Educational Resources* (HEDRES).

3.1 Summary of findings from TIMSS 2003

Before commenting on the comparative findings from fitting single level and multilevel models to the TIMSS 2003 data presented in **Appendix A3**, it is useful to examine the relationship between students' achievement scores for *Mathematics* and *Science*. Figure 3.1 presents the bivariate scatter plots and correlations (by student Gender) for the total Australian participating sample.

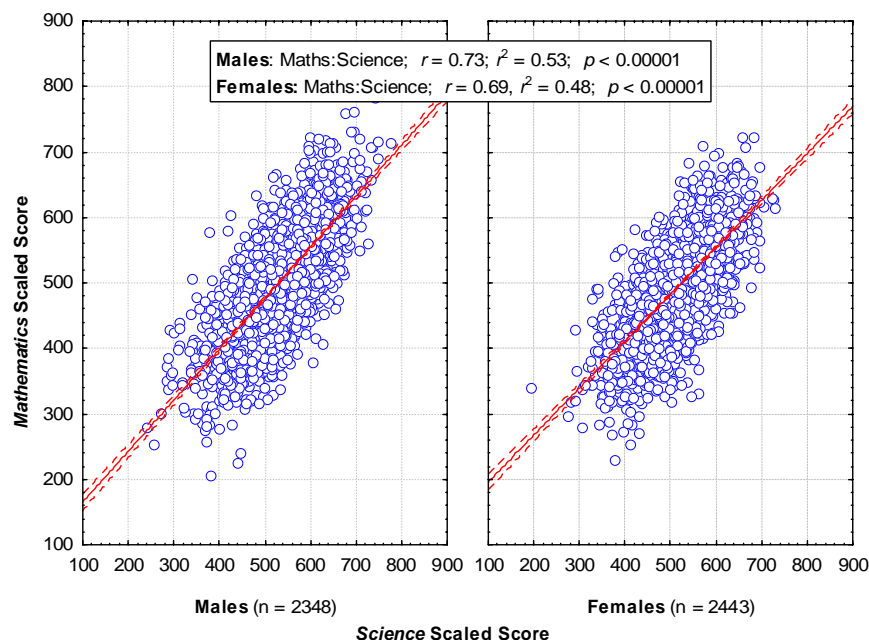


Figure 3.1 Relationship between TIMSS 2003 *Mathematics* and *Science* scaled scores for males and females, showing correlations (r) and coefficients of determination (r^2)

Comment: The relationships between students' *Mathematics* and *Science* scaled scores in TIMSS 2003 were strong and statistically significant accounting for 53% of the mutual variance in male students' *Mathematics* and *Science* scores, and 48% in female students' scores.

The State/Territory comparative findings of fitting single level and multilevel models to the TIMSS 2003 achievement data for students in their ninth year of formal schooling presented in **Appendix A3**, two findings are worthy of emphasis. First, based on the fitted single-levels models, after adjusting for students' background and 'intake' characteristics, there were significant variations in students' achievements across the States and Territories for both *Mathematics* (Fig. A3.1) and *Science* (Fig. A3.2).

Second, after adjusting for students' background and 'intake' characteristics, large and significant proportions of the residual variation were at the school-level for students' achievements in *Mathematics* (38.7%) and *Science* (27.7%). These results indicated significant between-school variations in students' TIMSS 2003 *Mathematics* and *Science* scores. Such

findings typically point to marked unevenness across schools in the quality of teaching and learning provision.

Concluding comments

The issues surrounding *school performance* and *educational effectiveness* are complex, multivariate and multidimensional. While Australia has much to be proud of its schools and the achievements of students within them, the findings summarised in this paper indicate considerable between-school variations within and between its eight States and Territories. Ultimately, *quality schooling* and *educational effectiveness* for all students is crucially dependent on the provision of **quality teaching** by competent teachers who are supported by capacity-building towards the maintenance of high teaching standards via strategic professional development at all levels of schooling (see: Darling-Hammond, 2000; Darling-Hammond & Bransford, 2005; Ingvarson, 2002, 2003).

Such outcomes, however, call for major reform requiring an investment in teacher quality that can then be used to change the ways in which students are taught and learn. Sadly, many educational reforms stop short of changing what happens beyond the classroom door, and thus fail to deliver improved teaching and learning outcomes for teachers and students, respectively. Rather, real reform directed at improving outcomes for all students – regardless of their backgrounds, ‘intake’ characteristics and residential locations – calls for substantial change in the quality of *teaching* and *learning* provision, but unless there is total commitment to teacher capacity-building, reform efforts soon falter.

It is important to note that the ‘myth’ of *school effectiveness* is grounded in a widespread failure to understand the fundamental distinction between *structure* and *function* in school education. Whereas a key *function* of schools is the provision of quality teaching and learning experiences that meet the developmental and learning needs of students is dependent on funding and organisational *structures* that support this function, the danger is a typical proclivity on the part of educational administrators to stress *structure* (e.g., single-sex schooling, class size, curriculum construction and reconstruction, etc.) at the expense of *function* (quality teaching and learning). Unfortunately, such emphases are indicative of a pervasive ignorance about what **really** matters in school education (i.e., quality teaching and learning), and the location of major sources of variation in students’ educational outcomes (i.e., the classroom).

It seems we need to be constantly reminded that schools and their structural arrangements are only as effective as the those responsible for making them work (school leaders and teachers) – in cooperation with those for whom they are charged and obligated to provide a professional service (students and parents). We also need to be reminded that the most valuable resources available to schools and the performance of their students are teachers. Thus, for the sake of our social and economic future – at the individual and national levels – we need to improve *school performance* by investing in *teacher* and *teaching quality*.

References

- ABS (1997). *Aspects of literacy: Assessed literacy skills*. Canberra, ACT: Australian Bureau of Statistics (4228.0). Available at: <http://www.abs.gov.au/Ausstats/abs@.nsf/0/887AE32D628DC922CA2568A900139365?Open>.
- ABS (2006). *Schools Australia 2005* (Cat. No: 4221.0). Canberra, ACT: Australian Bureau of Statistics.
- Beaton, A., Postlethwaite, T., Ross, K., Spearritt, D., & Wolf, R. (1999). *The benefits and limitations of international achievement studies*. Paris: International Institute for Educational Planning/UNESCO.
- Berlak, H. (1992). The need for a new science of assessment. In H. Berlak, F.M. Newmann, E. Adams, D.A. Archbald, T. Burgess, J. Raven, & T.A. Romberg (Eds.), *Toward a new science of educational testing and assessment* (pp. 139-180). Albany: State University of New York Press.
- Bosker, R.J., Creemers, B.P.M., & Scheerens, J. (Guest Editors) (1994). Conceptual and methodological advances in educational effectiveness research. *International Journal of Educational Research*, 21(2), 121-231.
- Browne, W., Healy, M., Cameron, B., & Charlton, C. (2005). *MLwiN software for multilevel analysis* (Version 2.02). Bristol, UK: Multilevel Models Project, University of Bristol (<http://multilevel.com>).
- Buckingham, J. (2003). *Schools in the Spotlight: School performance and public accountability* (CIS Policy Monograph 59). Sydney: The Centre for Independent Studies.
- Caldwell, B.J. (1993, February). *Liberating leaders in schools of the future*. Paper presented at the conference of the Victorian Primary Principals Association, Melbourne, February 15, 1993.
- Caldwell, B.J. (2003). Leadership for sustainable improvement for all students in all settings. In V. Zbar and T. Mackay (Eds.), *Leading the education debate: Selected papers from a decade of the IARTV Seminar Series* (pp. 322-342). Melbourne, VIC: Incorporated Association of Registered Teachers of Victoria (IARTV).
- Caldwell, B.J. (2004). *Re-imagining the self-managing school*. London: Specialist Schools Trust.
- Caldwell, B., Levacic, R. & Ross, K. (1999). The role of formula funding of schools in different educational policy contexts. In K. Ross & R. Levacic (Eds), *Needs-Based Resource Allocation in Education Via Formula Funding of Schools* (pp. 9-24.). Paris: UNESCO.
- Caldwell, B.J., & Spinks, J.M. (1988). *The self-managing school*. London: The Falmer Press.
- Caldwell, B.J., & Spinks, J.M. (1992). *Leading the self-managing school*. London: The Falmer Press.
- Chapman, J., Angus, L., Burke, G., & Wilkinson, V. (Eds.) (1991). *Improving the quality of Australian schools*. Australian Education Review No. 33. Hawthorn, VIC: Australian Council for Educational Research.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Cuttance, P. (2001). The impact of teaching on student learning. In K.J. Kennedy (Ed.), *Beyond the rhetoric: Building a teaching profession to support quality teaching* (pp. 35-55). Deakin West, ACT: Australian College of Education; College Year Book 2001.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1); available at: <http://epaa.asu.edu/epaa/v8n1>.
- Darling-Hammond, L., & Bransford, J. (Eds.) (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.
- DES (1984). *School standards and spending: Statistical analysis - a further appreciation*. London: Department of Education and Science.
- Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1), 1-32.
- Edmonds, R.R. (1979a). Some schools work and more can. *Social Policy*, 9, 28-32.
- Edmonds, R.R. (1979b). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-17.
- Edmonds, R.R. (1981). Making public schools effective. *Social Policy*, 12, 56-60.
- Forster, M. (2000). *A policy maker's guide to international achievement studies*. Camberwell, VIC: Australian Council for Educational Research.
- Forster, M. (2001a). Using assessment data. In M. Forster, G.N. Masters and K.J. Rowe, *Measuring learning outcomes: Options and challenges in evaluation and performance monitoring* (Revised edition, pp. 38-54). *Strategic Choices for Educational Reform; Module IV – Evaluation and Performance Monitoring*. Washington, DC: The World Bank Institute.

- Forster, M. (2001b). *A policy maker's guide to system-wide assessment programs*. Camberwell, VIC: Australian Council for Educational Research.
- Forster, M., Masters, G.N., & Rowe, K.J. (2001). Measuring learning outcomes: Options and challenges in evaluation and performance monitoring. *Strategic Choices for Educational Reform; Module IV – Evaluation and Performance Monitoring*. Washington, DC: The World Bank Institute.
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8, 369-395.
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: Scope and limitations. *British Educational Research Journal* 27(4), 433-442.
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*, 11, 319-330.
- Goldstein, H., & Healy, M.J.R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society, A*, 158, 175-177.
- Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance (with discussion). *Journal of the Royal Statistical Society, A*, 159, 385-443.
- Goldstein, H., & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society, A*, 159, 149-163.
- Goodlad, J. (1983). A study of schooling: Some implications for school improvement. *Phi Delta Kappan*, 64 (8), 552-558.
- Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.
- Greaney, V., & Kellaghan, T. (1996). *Monitoring the learning outcomes of education systems*. Washington, DC: The World Bank Institute.
- Hanusheck, E.A. (2004). *Some simple analytics of school quality*. NBER Working Paper 10229. Cambridge, MA: National Bureau of Economic Research Inc.
- Hanusheck, E.A. (2005a). *Economic outcomes and school quality*. Education Policy Series, Volume 4. Paris: International Institute for Educational Planning and International Academy of Education.
- Hanusheck, E.A. (2005b). The economics of school quality. *German Economic Review*, 6(3), 269-286.
- Hanusheck, E.A. (2005c). Why quality matters in education. *Finance and Development* (June 2005), 15-19.
- Hanusheck, E.A., & Jorgenson, D.W. (Eds.) (1996). *Improving America's Schools: The role of incentives*. Washington, DC: American Educational Research Association.
- Hanusheck, E.A., Kain, J.F., O'Brien, D.M., & Rivkin, S.G. (2005). *The market for teacher quality*. NBER Working Paper 11154. Cambridge, MA: National Bureau of Economic Research Inc.
- Hanusheck, E.A., & Raymond, M.E. (2004). *Does school accountability lead to improved student performance?* NBER Working Paper 10591. Cambridge, MA: National Bureau of Economic Research Inc.
- Hanushek, E.A., Rivkin, S.G., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Heckman, J.J. (2000). *Invest in the very young*. Working Paper, Harris Graduate School of Public Policy Studies. Available at: <http://www.HarrisSchool.uchicago.edu>.
- Heckman, J.J. (2005). *Lessons from the technology of skills formation*. National Bureau of Economic Research, Working Paper #1142, February 2005.
- Hill, P.W. (1995). Value added measures of achievement. *IARTV Seminar Series*, No. 44, May, 1995.
- Hill, P.W., & Crévola, C.A. (1999). The role of standards in educational reform for the 21st century. In D.D. Marsh (Ed.), *ASCD Year Book 1999: Preparing our schools for the 21st century* (pp. 117-142). Alexandria, VA: Association for Supervision and Curriculum Development.
- Hill, P.W., & Rowe, K.J. (1996). Multilevel modeling in school effectiveness research (Leading article). *School Effectiveness and School Improvement*, 7(1), 1-34.
- Hill, P.W., & Rowe, K.J. (1998). Modeling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement*, 9(3), 310-333.
- House of Commons Education and Skills Committee (2005, March). *Teaching children to read: Eighth Report of Session 2004-05*. London, UK: The Stationery Office Limited.

- Ingvarson, L. (2002). *Development of a National Standards Framework for the teaching profession*. An Issues paper prepared for the MCEETYA Taskforce on Teacher Quality and Educational Leadership. Camberwell, VIC: Australian Council for Educational Research.
- Ingvarson, L. (2003). A professional development system fit for a profession. In V. Zbar and T. Mackay (Eds.), *Leading the education debate: Selected papers from a decade of the IARTV Seminar Series* (pp. 391-408). Melbourne, VIC: Incorporated Association of Registered Teachers of Victoria (IARTV).
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gentis, H., Heynes, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.
- Kellaghan, T., Madaus, G.F., & Airasian, P.W. (1992). *The effects of standardized testing*. Boston: Kluwer-Nijhoff Publishing.
- Lacey, C., & Lawton, D. (1981). *Issues in evaluation and accountability*. London: Methuen.
- LaTrice-Hill, T. (2002). *No Child Left Behind Policy Brief: Teaching quality*. Denver, CO: Education Commission of the States. Available at: <http://www.ecs.org/clearinghouse/>.
- Lokan, J., Greenwood, L., & Cresswell, J. (2001). *15-up and counting, reading, writing, reasoning: how literate are Australia's students: the PISA 2000 survey of students' reading, mathematical and scientific skills*. Camberwell, VIC: Australian Council for Educational Research.
- Lyon, G.R. (2003). Reading disabilities: Why do some children have difficulty learning to read? What can be done about it? *Perspectives*, 29(2). Available from the International Dyslexia Association's website at: www.interdys.org.
- Manno, V.B. (1994). *Outcomes-based education: Miracle, cure or plague?* Hudson Institute Briefing Paper No. 165, June 1994.
- Marks, G.N. (2000). *The measurement of socioeconomic status and social class in the LSAY project*. Longitudinal Surveys of Australian Youth (LSAY), Technical Paper No. 14. Camberwell, VIC: Australian Council for Educational Research.
- Marks, G.N. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology*, 20(4), 483-505.
- Marks, G.N. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*, 48(1), 21-40.
- Marsh, H.W., Rowe, K.J., & Martin, A. (2002). PhD students' evaluations of research supervision: Issues, complexities and challenges in a nationwide Australian experiment in benchmarking universities (Leading article). *Journal of Higher Education*, 73(2), 313-348.
- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston, MA: International Association for the Evaluation of Education Achievement, Boston College.
- Masters, G. (1990). Improving the assessment of student outcomes. In J. Hewton (Ed.), *Performance indicators in education: What can they tell us?* (pp. 1-18). Brisbane, QLD: Australian Conference of Directors-General of Education.
- Masters, G.N. (1991). *Assessing achievement in Australian Schools*: A discussion paper commissioned by the Industry Education Forum. Hawthorn, VIC: Australian Council for Educational Research.
- Masters, G.N. (1994). *Setting and measuring performance standards for student achievement*. Paper presented at the conference "Public Investment in School Education: Costs and Outcomes", sponsored by The Schools Council and The Centre for Economic Policy Research, Australian National University, Canberra, March 17, 1994.
- Masters, G.N. (2004). *Continuity and growth: Key considerations in educational improvement and accountability*. Background paper to keynote address Presented at the joint Australian College of Educators and Australian Council for Educational Leaders national conference, Perth, Western Australia, 6-9 October 2004.
- Masters, G.N., & Forster, M. (1997a). *Mapping literacy achievement: Results of the 1996 National School English Literacy Survey*. A report on behalf of the Management Committee for the National School English Literacy Survey. Canberra, ACT: Commonwealth Department of Employment, Education, Training and Youth Affairs.
- Masters, G.N., & Forster, M. (1997b). *Literacy standards in Australia*. Canberra, ACT: Commonwealth of Australia.

- MCEETYA (2005). *National report on schooling in Australia, Preliminary paper: National Benchmark results for Reading, Writing and Numeracy in Years 3, 5 and 7*. Canberra, ACT: Ministerial Council on Education, Employment, Training and Youth Affairs.
- McGaw, B. (1991). Monitoring education systems. In J. Chapman, L. Angus, G. Burke, & V. Wilkinson (Eds.), *Improving the quality of Australian schools*. Australian Education Review No. 33 (pp. 134-139). Hawthorn, VIC: Australian Council for Educational Research.
- Mortimore, P. (1991). School effectiveness research: Which way at the crossroads? *School Effectiveness and School Improvement*, 2 (3), 213-229.
- Mortimore, P. (1992). Issues in school effectiveness. In D. Reynolds & P. Cuttance (Eds.), *School effectiveness: Research, policy and practice* (pp. 154-163). London: Cassell.
- Mortimore, P. (1995). *Effective schools: Current impact and future potential*. Director's Inaugural Lecture. Institute of Education, University of London.
- Mortimore, P. (2001). Globalisation, effectiveness and improvement. *School Effectiveness and School Improvement*, 12 (2), 229-249.
- Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice*. London: Paul Chapman Publishing.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston, MA: International Association for the Evaluation of Education Achievement, Boston College.
- Murphy, P., Greaney, V., Lockheed, M., & Rojas, C. (Eds.) (1996). *National Assessments: Testing the system*. Washington, DC: The World Bank Institute.
- NCEE-National Centre on Education and the Economy (1997). *New standards for performance*. Washington, DC: Author.
- Nisbet, J. (1993). Introduction. In *OECD - Curriculum reform: Assessment in question* (pp. 25-38). Paris: Organisation for Economic Cooperation and Development.
- Nuttall, D.L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, 13(7), 769-776.
- OECD (1983). *Compulsory schooling in a changing world*. Paris: Organisation for Economic Cooperation and Development.
- OECD (1986). *Education and training for manpower development*. Paris: Organisation for Economic Cooperation and Development.
- OECD (1989). *Schools and quality: An international report*. Paris: Organisation for Economic Cooperation and Development.
- OECD (1993). *Curriculum reform: Assessment in question*. Paris: Organisation for Economic Cooperation and Development.
- OECD, (1995). *Indicators of education systems: Measuring the quality of schools*. Paris: Organization for Economic Cooperation and Development.
- OECD (2001). *Teachers for tomorrow's schools: Analysis of the World Education Indicators, 2001 edition*. Paris: Organisation for Economic Cooperation and Development and UNESCO Institute for Statistics.
- OECD (2002). Improving both quality and equity: Insights from PISA 2000. In *Education Policy Analysis 2002* (pp. 35-63). Paris: OECD.
- OECD (2003). *The PISA 2003 assessment framework: Reading, mathematical and scientific literacy*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2005a). *Education at a glance: OECD indicators 2005*. Paris: Organization for Economic Cooperation and Development.
- OECD (2005b). *Teachers matter: Attracting, developing and retaining effective teachers*. Paris: Organization for Economic Cooperation and Development.
- Plomp, T. (1999). *The relevance of IEA type international comparative assessments of educational achievement*. Paper presented at the 40th General Assembly of IEA Oslo, 30 August 1999.
- Podgursky, M., Monroe, R., & Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review*, 23(5), 507-518.
- Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement*, 4(1), 37-58.

- Rothman, S. (2002). Achievement in literacy and numeracy by Australian 14 year-olds, 1975-1998. *Longitudinal Surveys of Australian Youth, Research Report No. 29*. Camberwell, VIC: Australian Council for Educational Research. Available for download in PDF format at: <http://www.acer.edu.au/research/lsey/reports/LSAY29.pdf>.
- Rowe, K.J. (2000). Assessment, league tables and school effectiveness: Consider the issues and let's get real! *Journal of Educational Enquiry*, 1(1), 72-97.
- Rowe, K.J. (2001a). Educational performance indicators. In M. Forster, G.N. Masters and K.J. Rowe, Measuring learning outcomes: Options and challenges in evaluation and performance monitoring (Revised edition, pp. 2-20). *Strategic Choices for Educational Reform; Module IV – Evaluation and Performance Monitoring*. Washington, DC: The World Bank Institute.
- Rowe, K.J. (2001b, October). *Responsible and irresponsible uses of 'value-added' measures in educational assessment and reporting*. Invited keynote address presented at the Sixth National Roundtable on Assessment and Reporting, Hobart, October 24-26, 2001.
- Rowe, K.J. (2003). *The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling*. Background paper to keynote address presented at the 2003 ACER Research Conference, Carlton Crest Hotel, Melbourne, 19-21 October 2003. Available in PDF format at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2004a). *Analysing and reporting performance indicator data: 'Caress' the data and user beware!* Background paper to invited address presented at the 2004 Public Sector Performance & Reporting Conference (under the auspices of the International Institute for Research – IIR), Sydney, 19-22 April 2004. Available at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2004b). *The importance of teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision – implications of findings from the international and Australian evidence-based research*. Background paper to invited address presented at the *Making Schools Better* summit conference, Melbourne Business School, the University of Melbourne, 26-27 August 2004. Available in PDF format at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2005a). *Performance indicators of quality schooling: Their nature, construction and use*. Background paper of presentation to senior educational administrators, the Philippines, 21 November, 2005. Camberwell, VIC: Australian Council for Educational Research.
- Rowe, K.J. (2005b). Evidence for the kinds of feedback data that support both student and teacher learning. *Research Conference 2005 Proceedings* (pp. 131-146). Camberwell, VIC: Australian Council for Educational Research [ISBN 0-86431-684-4]. Available for download in PDF format at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (Chair) (2005c). *Teaching reading: Report and recommendations*. Report of the Committee for the National Inquiry into the Teaching of Literacy. Canberra, ACT: Australian Government Department of Education, Science and Training. Available at: www.dest.gov.au/nitl/report.htm.
- Rowe, K.J., & Hill, P.W. (1996). Assessing, recording and reporting students' educational progress: The case for 'Subject Profiles'. *Assessment in Education*, 3(3), 309-352.
- Rowe, K.J., & Hill, P.W. (1998). Modeling educational effectiveness in classrooms: The use of multilevel structural equations to model students' progress. *Educational Research and Evaluation*, 4(4), 307-347.
- Rowe, K.J., & Lievesley, D. (2002, April). *Constructing and using educational performance indicators*. Background paper to keynote address and workshops presented at the inaugural Asia-Pacific Educational Research Association (APERA) regional conference, ACER, Melbourne, April 16-19, 2002; available at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J., & Rowe, K.S. (1999). Investigating the relationship between students' attentive-inattentive behaviors in the classroom and their literacy progress. *International Journal of Educational Research*, 31(1/2), 1-138 (Whole Issue). Elsevier Science, Pergamon Press.
- Rowe, K.J., & Stephanou, A. (2001, October). *'Value-added' educational performance indicators using measurement and multilevel modelling*. Invited workshop presented at the Sixth National Roundtable on Assessment and Reporting, Hobart, October 24-26, 2001.
- Rowe, K.J., & Stephanou, A. (2003). *Performance audit of literacy standards in Victorian Government schools, 1996-2002*. A consultancy report to the Victorian Auditor General's Office. Melbourne, VIC: Australian Council for Educational Research.

- Rowe, K.J., Turner, R., & Lane, K. (2002). Performance feedback to schools of students' Year 12 assessments: The *VCE Data Project*. In A.J. Visscher and R. Coe (Eds.), *School improvement through performance feedback* (pp. 163-190). Lisse, The Netherlands: Swetz & Zeitlinger.
- Rutter, M., Maughan, B., Mortimer, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Somerset: Open Books.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- StatSoft, Inc. (2005). *STATISTICA data analysis software system* (Version 7.2). Tulsa, AR: StatSoft Incorporated (www.statsoft.com).
- Stephanou, A., & Rowe, K.J. (2002). *WAMSE-English 2001: Multilevel modelling*. A report to the Department of Education, Western Australia, of key measurement and multilevel modelling findings related to the *Reading and Viewing* achievements of students in Years 3, 7 and 10. Camberwell, VIC: Australian Council for Educational Research.
- Summit of the America's – Line 2: Educational Assessment, Brasilia*, March 12-14, 2002. INEP.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years (Leading article). *School Effectiveness and School Improvement*, 8, 169-197.
- Thomson, S., Cresswell, J., & De Bortoli, L. (2004). *Facing the future: A focus on mathematical literacy among Australian 15-year-old students in PISA 2003*. Camberwell, VIC: Australian Council for Educational Research.
- Thomson, S., Rowe, K.J., Underwood, C., & Peck, R. (2005). *Numeracy in the early years: Project Good Start*. Final report to the Australian Government Department of Education, Science and Training. Camberwell, VIC: Australian Council for Educational Research. Available at: http://www.dest.gov.au/sectors/school_education/publications_resources/profiles/goodstart.htm#publication; and at: <http://www.acer.edu.au/research/projects/goodstart/documents/GoodstartFinalReport.pdf>.
- Tucker, M.S., & Coddling, J.B. (1998). *Standards for our schools: How to set them, measure them and reach them*. San Francisco, CA: Jossey-Bass.
- US Department of Education (2002). *No Child Left Behind: A desktop reference*. Washington, DC: Author. Available at: www.ed.gov/offices/OESE/reference.
- Visscher, A.J., & Coe, R. (Eds.) (2002). *School improvement through performance feedback*. Lisse, The Netherlands: Swetz & Zeitlinger.
- Visscher, A., Karsten, S., de Jong, T., & Bosker, R. (2000). Evidence on the intended and unintended effects of publishing school performance indicators. *Evaluation and Research in Education*, 14, 254-267.
- Watson, L. (1996). Public accountability or fiscal control? Benchmarks of performance in Australian schooling. *Australian Journal of Education*, 40, 104-123.
- Willms, J.D. (2000). Monitoring school performance for standards-based reform. *Evaluation and Research in Education*, 14, 237-253.
- Wyatt, T. (1994). Education indicators: A review of the literature. In *Making education count: Developing and using international indicators* (pp. 99-121). Paris: OECD.
- Zammit, S., Routitsky, A., & Greenwood, L. (2002). *Mathematics and Science achievement of junior secondary school students in Australia*. Camberwell, VIC: Australian Council for Educational Research.

Appendix A1

State/Territory Comparisons of Student Achievement

Benchmark data for Literacy and Numeracy

For the purposes of this Appendix, the 'benchmark' profiles by calendar year for the State or Territory with the highest average percentage of students above the 'benchmark' are bolded.

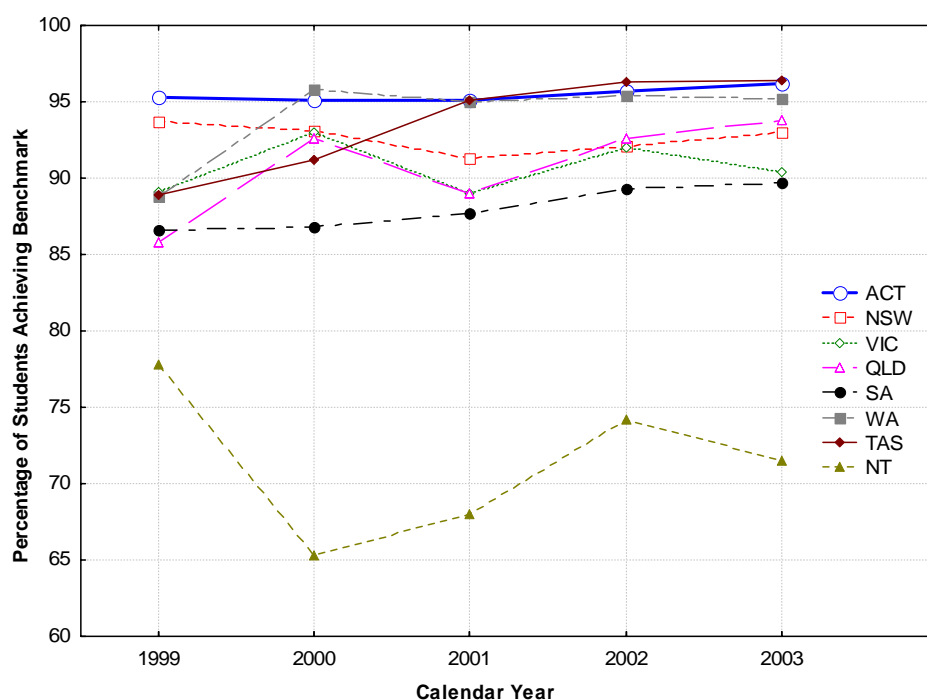


Figure A1.1 Grade 3 *Reading* benchmark profiles by State/Territory, 1999-2003

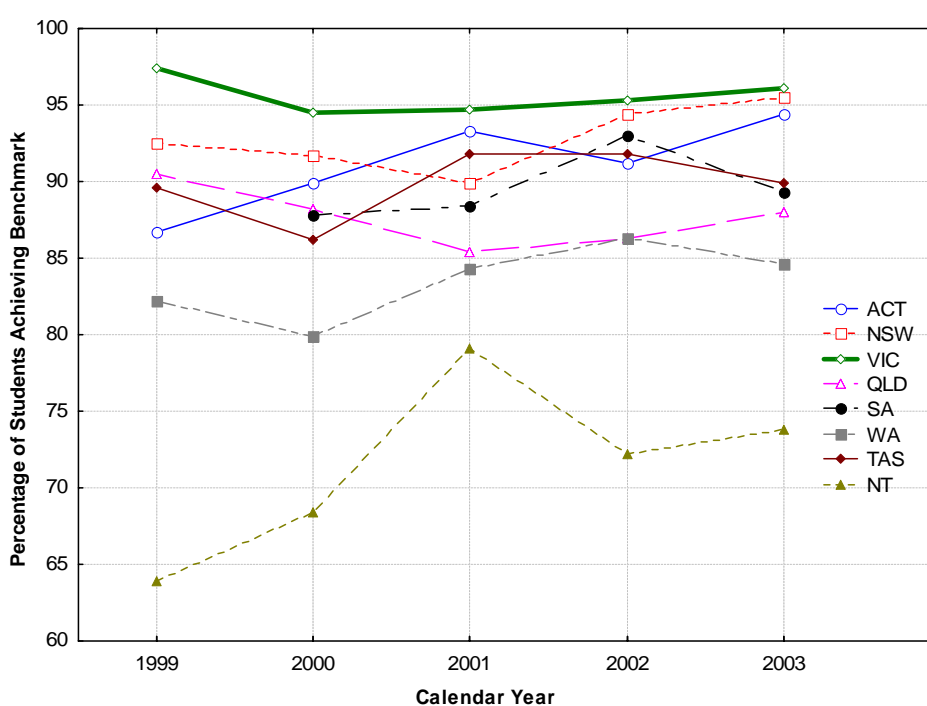


Figure A1.2 Grade 3 *Writing* benchmark profiles by State/Territory, 1999-2003

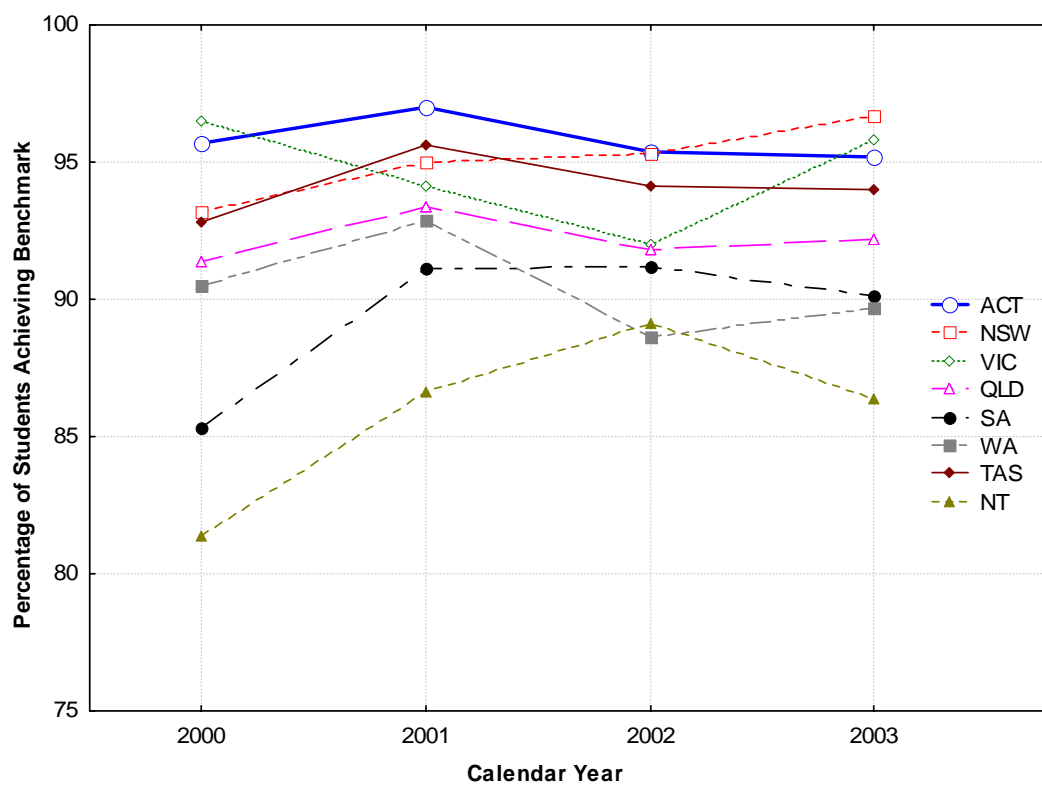


Figure A1.3 Grade 3 Numeracy benchmark profiles by State/Territory, 2000-2003

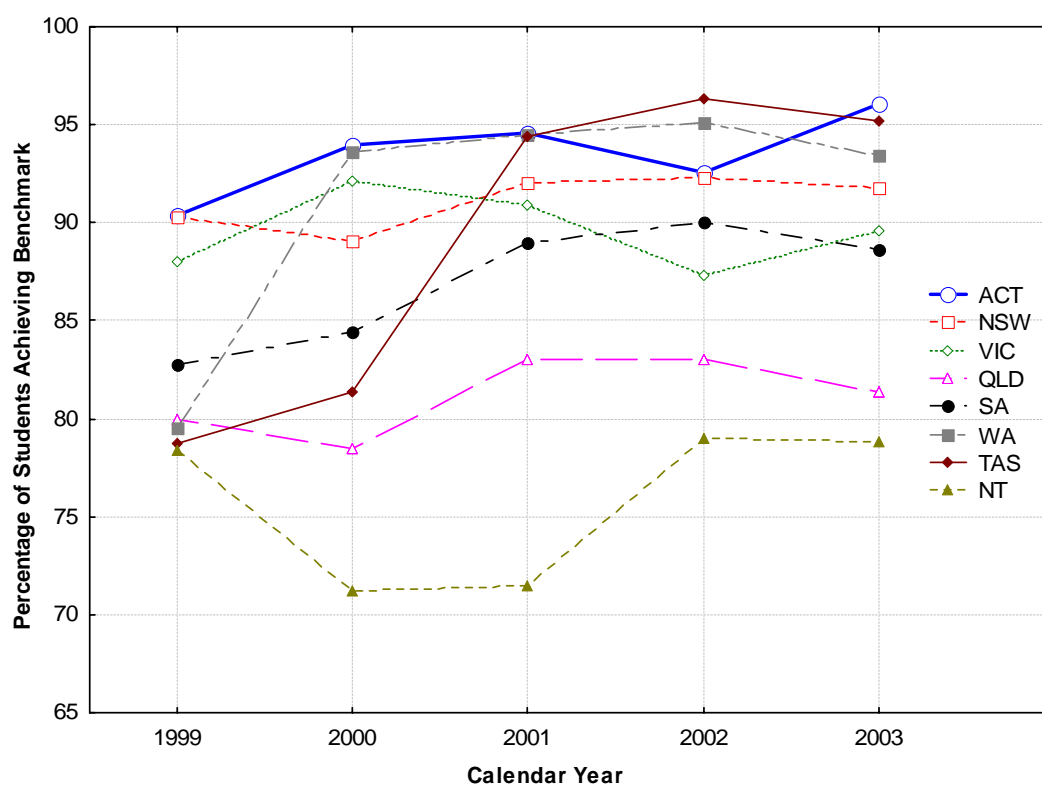


Figure A1.4 Grade 5 Reading benchmark profiles by State/Territory, 1999-2003

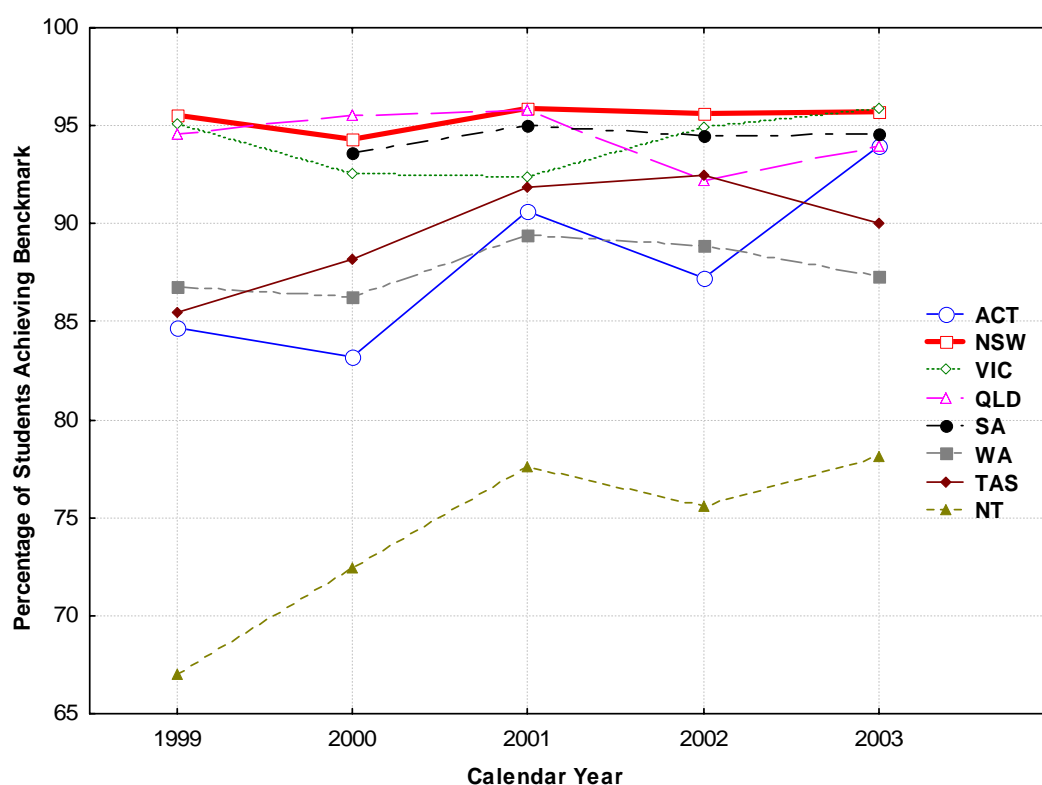


Figure A1.5 Grade 5 *Writing* benchmark profiles by State/Territory, 1999-2003

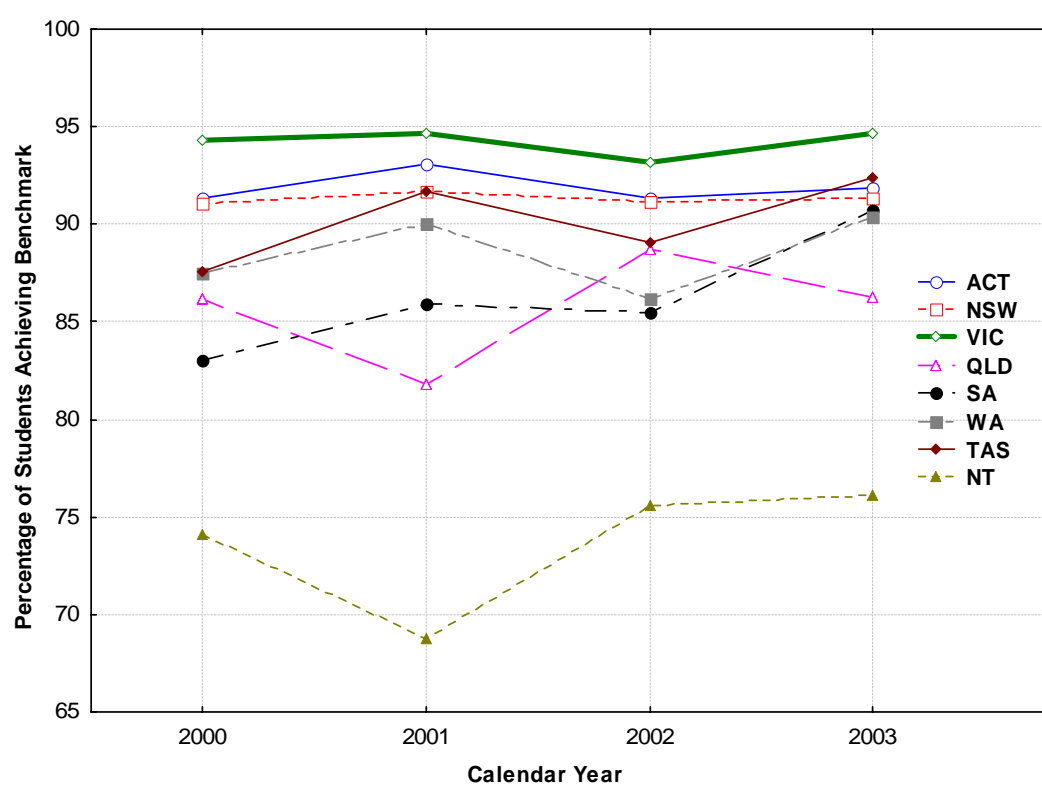


Figure A1.6 Grade 5 *Numeracy* benchmark profiles by State/Territory, 2000-2003

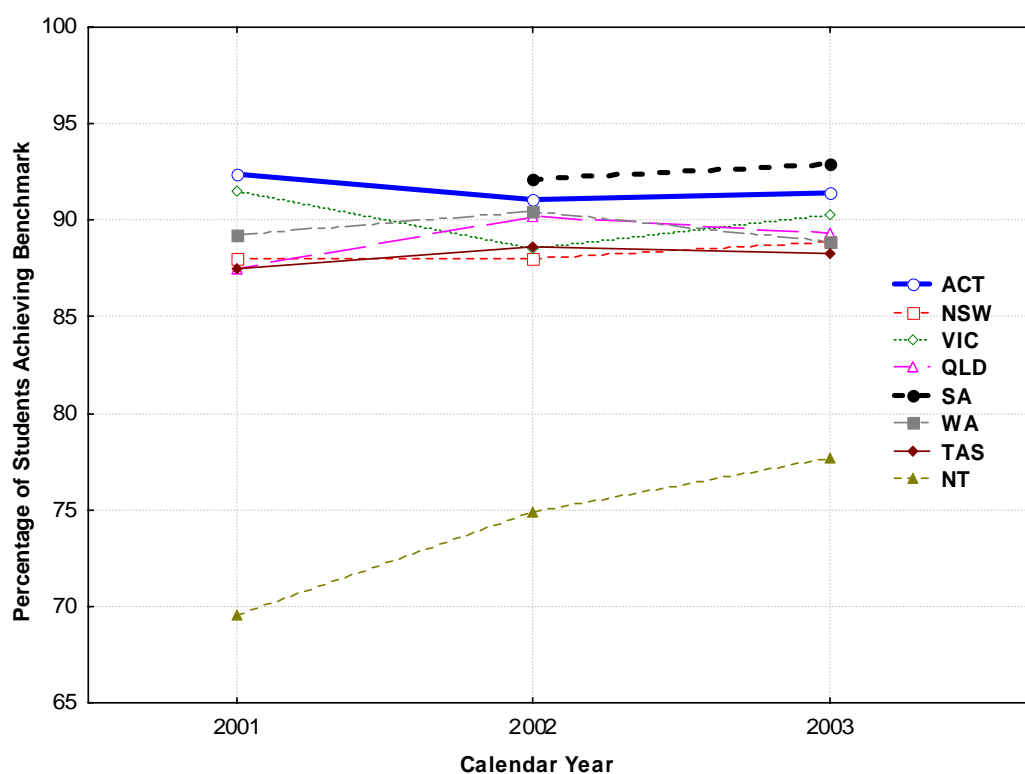


Figure A1.7 Grade 7 Reading benchmark profiles by State/Territory, 2001-2003

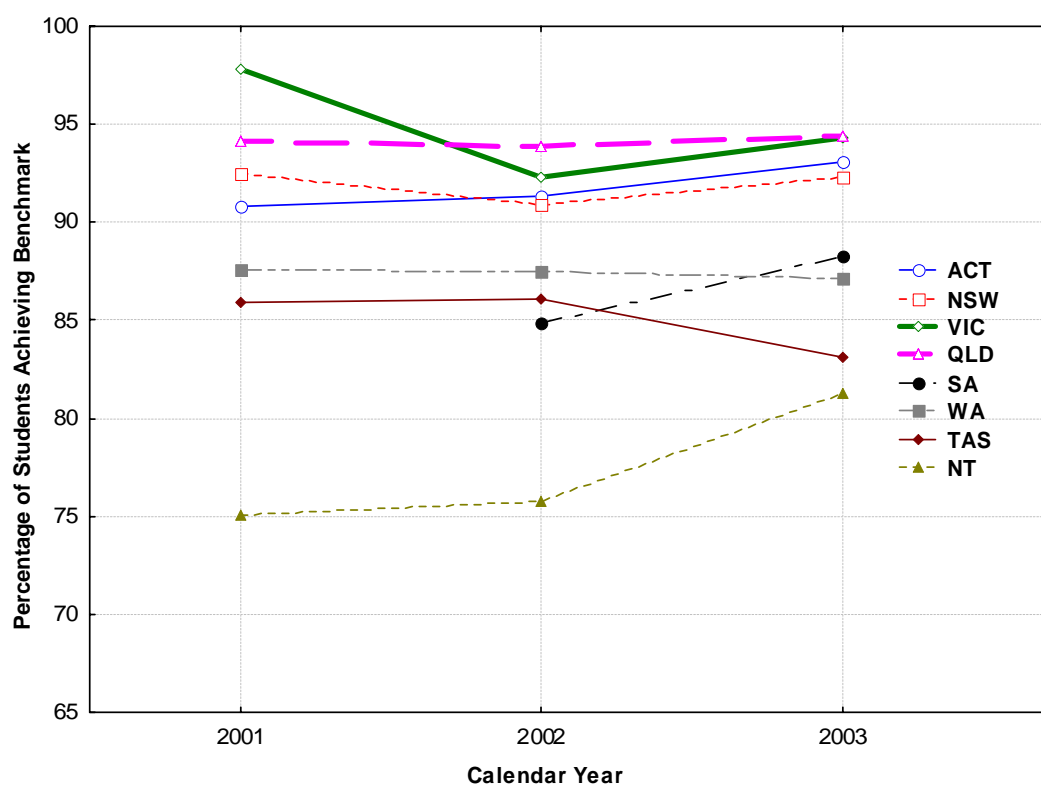


Figure A1.8 Grade 7 Writing benchmark profiles by State/Territory, 2001-2003

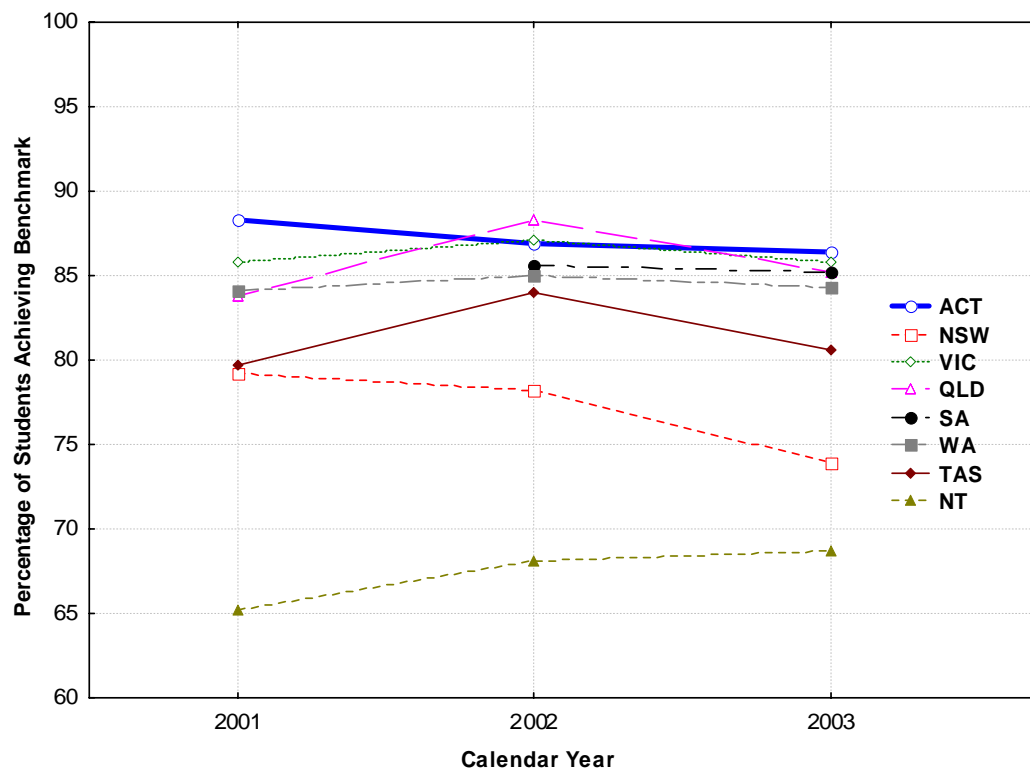


Figure A1.9 Grade 7 Numeracy benchmark profiles by State/Territory, 2001-2003

Appendix A2

State/Territory Comparisons of Students' Achievements in PISA

PISA 2000: Single-level multivariate analyses

The following State/Territory comparisons derive from fitting single-level multivariate models to the PISA 2000 *Reading Literacy* and *Mathematical Literacy* achievement scores obtained from 15 year-old students – randomly selected within a stratified sample of Australian government, Catholic and independent schools (for specific details and definitions of achievement and presage variables, see Lokan, Greenwood & Cresswell, 2001). For present purposes, the achievement scores are adjusted for *State/Territory* location, *Gender*, *Age*, *Grade*, *SES*, and *Home Educational Resources* (HEDRES). It is important that these findings from fitting single-level models to the student achievement data (as summarised below) are not over-interpreted, since such models ignore the inherent hierarchical structure of the data; i.e., 5176 students' *Reading Literacy* and 2859 students' *Mathematical Literacy* achievement scores (level-1), clustered within 231 schools (level-2) and 8 States/Territories (level-3). Nonetheless, the graphical presentations provide initial indications of the relative performances of male and female students within each of the 8 States and Territories.

PISA 2000 Reading Literacy

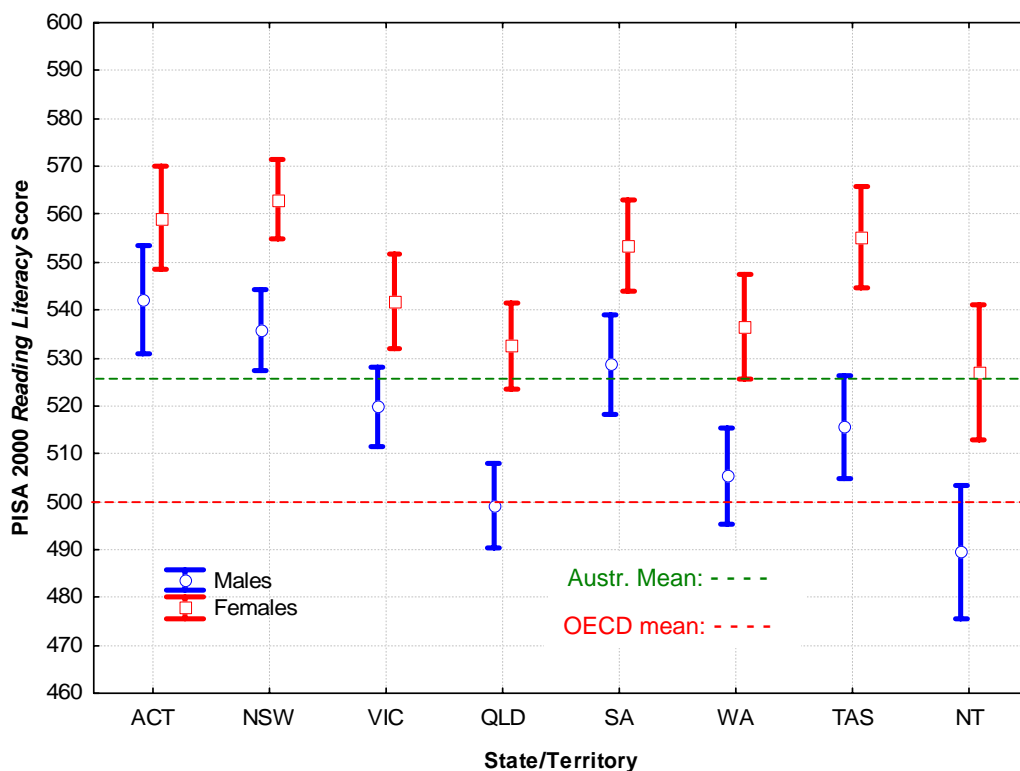


Figure A2.1 Adjusted mean-point estimates for PISA 2000 *Reading Literacy* scores bounded by 95% confidence intervals for students, by State/Territory and Gender

State/Territory effect: $F(7,4867) = 14.8, p < 0.000001$
 Gender effect: $F(1,4867) = 120.5, p < 0.000001$
 Age effect: $F(1,4867) = 0.47, p = 0.495$ (n.s.)
 Grade effect: $F(1,4867) = 242.3, p < 0.000001$
 SES effect: $F(1,4867) = 377.7, p < 0.000001$
 HEDRES effect: $F(1,4867) = 269.8, p < 0.000001$
 State/Territory \times Gender effect: $F(7,4867) = 1.01, p = 0.425$ (n.s.)

Comment: Whereas the findings from fitting this single-level model to the PISA 2000 *Reading Literacy* score data indicate significant differences in students' achievements between the States and Territories, these differences are misleading because their standard errors (and hence, confidence intervals) are based on the total number of participating Australian students rather than the 8 States and Territories. Compared with the findings obtained from fitting a multilevel model to these data (see Fig. A2.4), this is a classic case of a Type I error (i.e., unjustifiably claiming a 'difference').

With the exception of the effects for *Age* and *State/Territory* \times *Gender* interaction, all other effects of the fitted variables were statistically significant, the most notable of which were *Gender* (in favour of females), *SES* (in favour of higher *SES*), *Grade* (in favour of higher Grades), and *Home Educational Resources* (HEDRES – in favour of higher HEDRES).

PISA 2000 *Mathematical Literacy* (Australia)

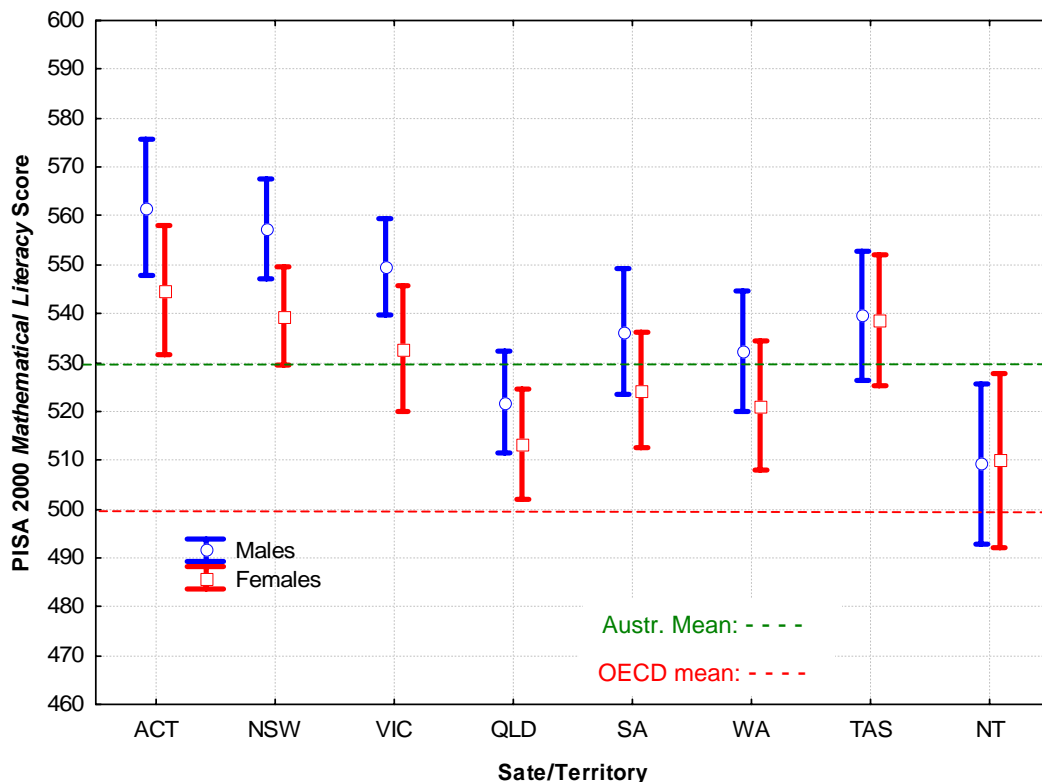


Figure A2.2 Adjusted mean-point estimates for *Mathematical Literacy* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,2713) = 9.2, p < 0.000001$

Gender effect: $F(1,2713) = 10.2, p < 0.01$

Grade effect: $F(1,2713) = 206.4, p < 0.000001$

SES effect: $F(1,2713) = 306.3, p < 0.000001$

HEDRES effect: $F(1,2713) = 49.5, p < 0.000001$

State/Territory \times Gender effect: $F(7,2713) = 0.55, p = 0.799$ (n.s.)

Comment: Again, the findings from fitting this single-level multivariate model to the PISA 2000 *Mathematical Literacy* score data indicate significant differences between the States and Territories. With the exception of the effect for *State/Territory* \times *Gender* interaction, all other effects of the fitted variables were statistically significant, the most notable of which were: *SES* (in favour of higher *SES*), *Grade* (in favour of higher Grade membership), *Home Educational Resources* (HEDRES – in favour of higher HEDRES), and *Gender* (in favour of males). It is interesting to note that the PISA 2000 *Mathematical Literacy* mean scores for males and females

within each of the States and Territories were not significantly different, including the ACT, since the corresponding confidence intervals overlap. However, these differences are misleading because their standard errors are based on the total number of participating Australian students rather than within each of the 8 States and Territories.

Multilevel analyses of PISA 2000 data

Following are results of the fitted baseline Variance Components model for 5176 and 2859 15 year-old students (i) in 231 schools (j) drawn from 8 Australian States/Territories (k) – based on students' normalised¹² scaled scores for PISA 2000 *Reading Literacy* and *Mathematical Literacy*, respectively. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

PISA 2000 *Reading Literacy*

$$\text{Reading}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.034(0.057) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.018(0.013)] \quad \text{Between-State/Territory res. var: 1.8\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.187(0.021)] \quad \text{Between-schools res. var: 18.7\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.797(0.016)] \quad \text{Within-schools res. var: 79.5\%}$$

Comment: Around the normalised grand mean for Australia (-0.034), the residual variance in students' PISA 2000 *Reading Literacy* achievement scores between States and Territories were **not** statistically significant. Unadjusted residual plots at the State/Territory-level illustrate these results are presented in Figure A2.3 below.

From Figure A2.3, with the exception of the NT (below), the 'uncertainty' intervals¹³ around the unadjusted means for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level in students' PISA 2000 *Reading Literacy* scores – accounting for a mere 1.8% of the residual variance. As indicated by Rowe (2004a), raw, unadjusted, comparative performances of the group kind presented graphically in Figure A2.3 have little utility other than to delude public sector administrators

¹² Given that the continuous response and explanatory variables of subsequent interest here are measured in different metrics (i.e., *Reading*, *Age*, *Grade SES*, *HEDRES* and *ScAvSES*), such variables should be recomputed as Normal scores, namely, as 'normal equivalent deviates' (NEDs) under the Normal distribution, for two reasons: (a) to ensure that such variables are 'measured' on a common metric, and (b) to assist in the comparative interpretation of 'effect sizes' of the fitted explanatory variables – expressed as standard deviation units (SDs).

¹³ Rather than referring to these intervals as **uncertainty intervals** (UIs), it is more common to refer to them as **confidence intervals** (CIs). Note that 95% *confidence intervals* for a statistic (a mean point-estimate for each State/Territory in this case – \bar{x}_s) are calculated from: $\bar{x}_s \pm 1.96 \times$ the State's standard error (i.e., the State's standard deviation divided by the square root of the State's cohort size – $\sigma_s / \sqrt{n_s}$). These intervals imply that we can be '95% confident' that the estimate of a State's mean lies between these upper and lower limits. However, in the present context of making comparative judgments about the relative performance of States/Territories and schools, these limits are more properly referred to as **uncertainty intervals**. That is, when the intervals for two or more States/Territories and/or schools overlap, there is **no certainty** that their relative performance differs significantly. For a presentation and discussion of the relevant conceptual and technical issues, see: Goldstein and Healy (1995).

(and politicians) by engendering greater or lesser degrees of ‘self-satisfaction’, ‘indifference’ or ‘despair’. Nonetheless, whereas the estimates obtained from fitting a variance-components model to the data are not of particular interest (*per se*), they provide a useful baseline from which to compare findings from fitting more ‘responsible’ models.

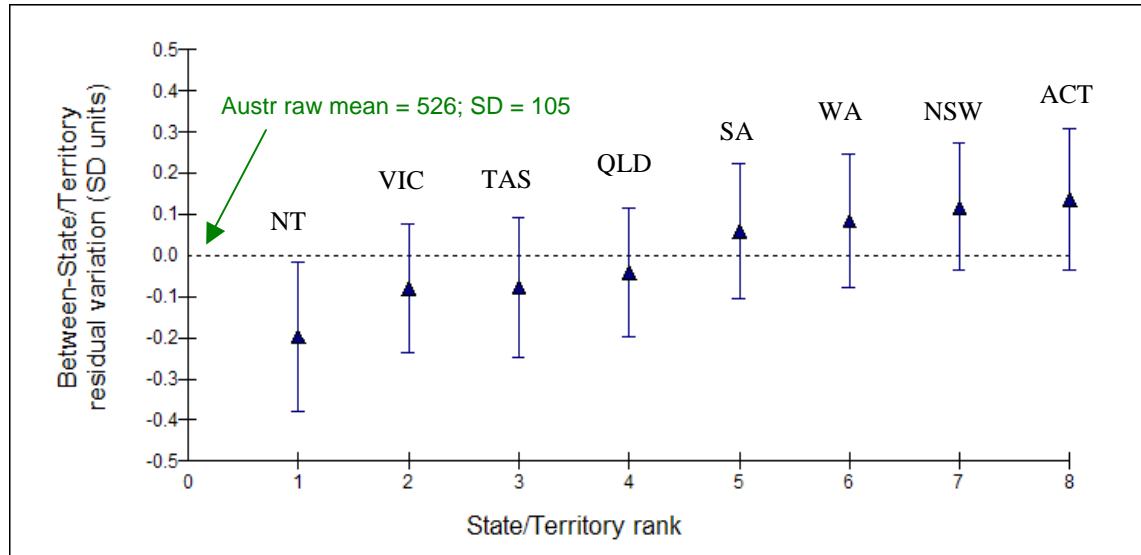


Figure A2.3 Ranked State/Territory-level raw residuals for PISA 2000 *Reading Literacy* scores, showing mean-point estimates bounded by 95% ‘uncertainty’ intervals (see footnote 4)

In the following model, adjustments are made for the ‘intake’ variables of *Gender*, *Age*, *Grade*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average SES at the school-level (i.e., *ScAvSES* – to estimate the within-school average ‘cohort effect’ of SES, over-and-above that operating at the individual student-level).

The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses. [Note: statistical significance at the $p < 0.05$ level is indicated when parameter estimates are at least twice the magnitude of their corresponding standard errors].

$$\text{Reading}_{ijk} = \beta_{0ijk} \text{Cons} + 0.273(0.025) \text{Gender}_{ijk} + -0.002(0.004) \text{Age}_{ijk} + 0.308(0.021) \text{Grade}_{ijk} + 0.169(0.014) \text{SES}_{ijk} + 0.190(0.013) \text{HEDRES}_{ijk} + 0.210(0.019) \text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = 0.257(0.760) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.012(0.007)] \quad \text{Between-State/Territory res. var: 1.7\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.038(0.007)] \quad \text{Between-schools res. var: 5.5\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.640(0.013)] \quad \text{Within-schools res. var: 92.8\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of females), *Grade* (in favour of higher Grade membership), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES). That is, these variables were significant predictors of Australian students’ PISA 2000 *Reading Literacy* achievement scores, while the effect of *Age* was small and not significant.

Together, all six fitted variables accounted for 31.1% of the variance in students' achievement scores, with an insignificant 1.7% of the residual variance at the State/Territory-level, and a small but significant 5.5% of the residual variance due to variation between schools. As expected, the bulk of the residual variance was at the student-level (i.e., 92.8%). Mean-adjusted residual plots at the State/Territory-level illustrate these results in Figure A2.4.

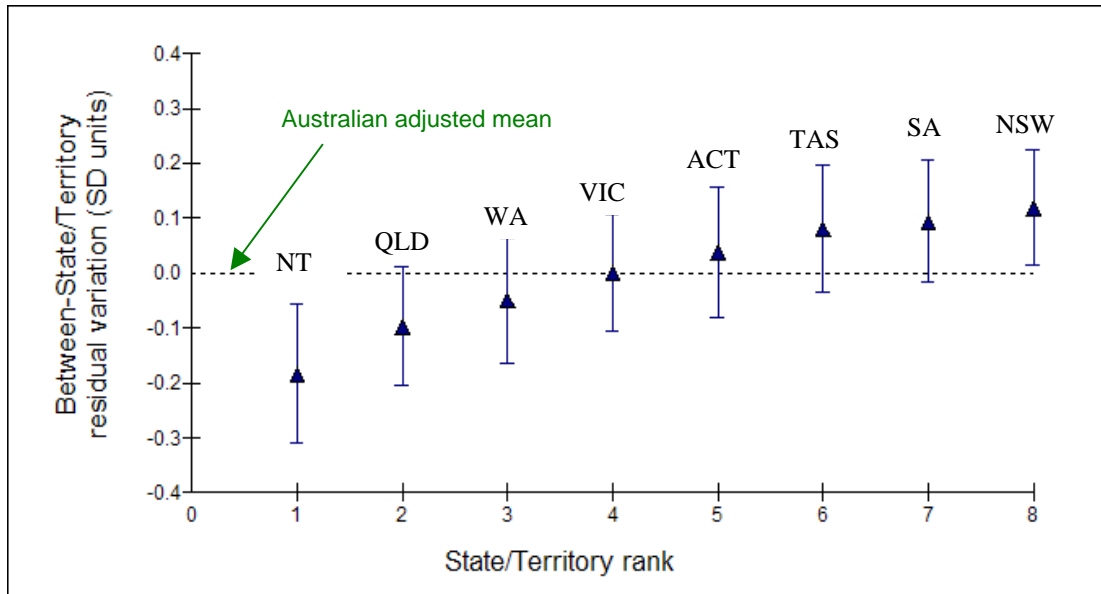


Figure A2.4 Plot of ranked residuals at the State-level, showing adjusted mean-point PISA 2000 *Reading Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: With the exception of NT (below) and NSW (above), the 95% 'uncertainty' intervals around the adjusted means, the intervals for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., a small 1.7% of the residual variance in students' PISA 2000 *Reading Literacy* scores). Indeed, there were no significant differences between the adjusted mean performances of students located in WA, VIC, ACT, TAS, SA and NSW schools. Nonetheless the performance of students in NT and QLD schools were less than satisfactory.

Although these results are of minor interest, they mask the variation between-schools at the national level, as well as between-school variation within each of the States and Territories separately. Following in Figure A2.5 is a plot of ranked residuals for 231 Australian schools from a multilevel analysis of residuals for performance in PISA 2000 *Reading Literacy*, showing adjusted mean-point estimates bounded by 95% 'uncertainty' intervals.

Comment: From Figure A2.5, the 'uncertainty' intervals around the adjusted means for approximately 90% the 231 schools all overlap the national mean (zero) – indicating non-significant differences in students' achievements between these schools (i.e., a small but significant 5.5% of the residual variance in students' PISA 2000 *Reading Literacy* scores). Nevertheless, the remaining 10% of schools yielded student performances either significantly above or below the national mean – again indicating differences in the quality of teaching and learning provision among Australian schools.

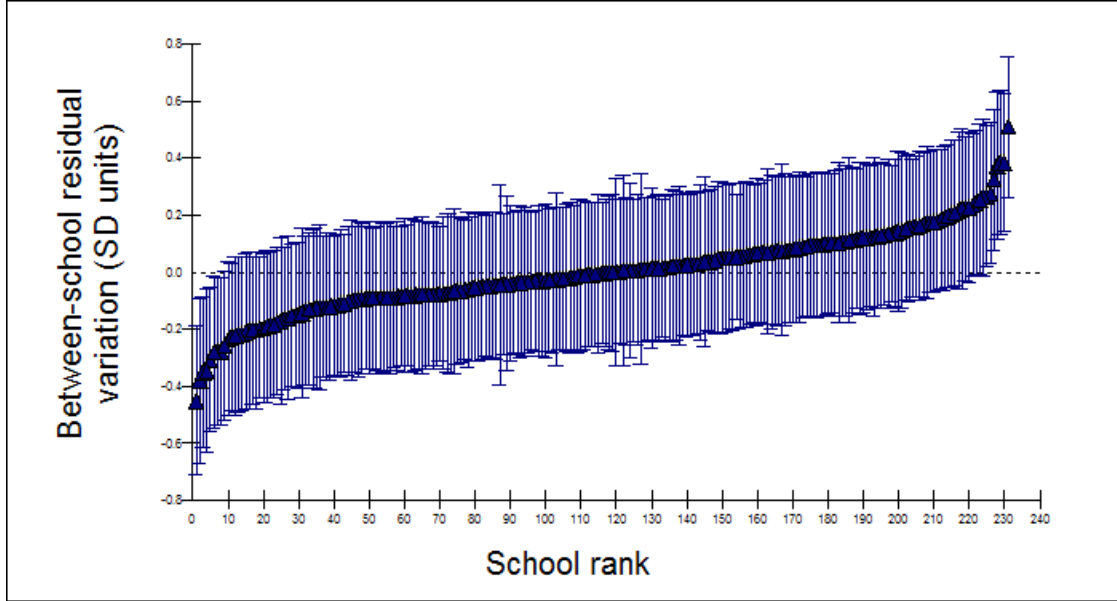


Figure A2.5 Plot of ranked residuals for 231 schools, showing adjusted mean-point PISA 2000 *Reading Literacy* score estimates, bounded by 95% ‘uncertainty’ intervals

PISA 2000 *Mathematical Literacy*

Following are results of the fitted baseline Variance Components model for 2859 15 year-old students (i) in 231 schools (j) drawn from 8 Australian States/Territories (k) – based on students’ normalised scaled scores for PISA 2000 *Mathematical Literacy*. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

$$\text{Mathematics}_{ijk} = \beta_{0ijk} \text{cons}$$

$$\beta_{0ijk} = -0.032(0.058) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.018(0.013) \end{bmatrix} \text{ Between-State/Territory res. var: 1.8\% (n.s.)}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.181(0.024) \end{bmatrix} \text{ Between-schools res. var: 18.0\%}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.804(0.022) \end{bmatrix} \text{ Within-schools res. var: 80.2\%}$$

Comment: Around the normalised grand mean (-0.032), the residual variance in students’ PISA 2000 *Mathematical Literacy* achievement scores between States and Territories were **not** statistically significant. Unadjusted residual plots at the State/Territory-level shown in Figure A2.6 illustrate these results as follows.

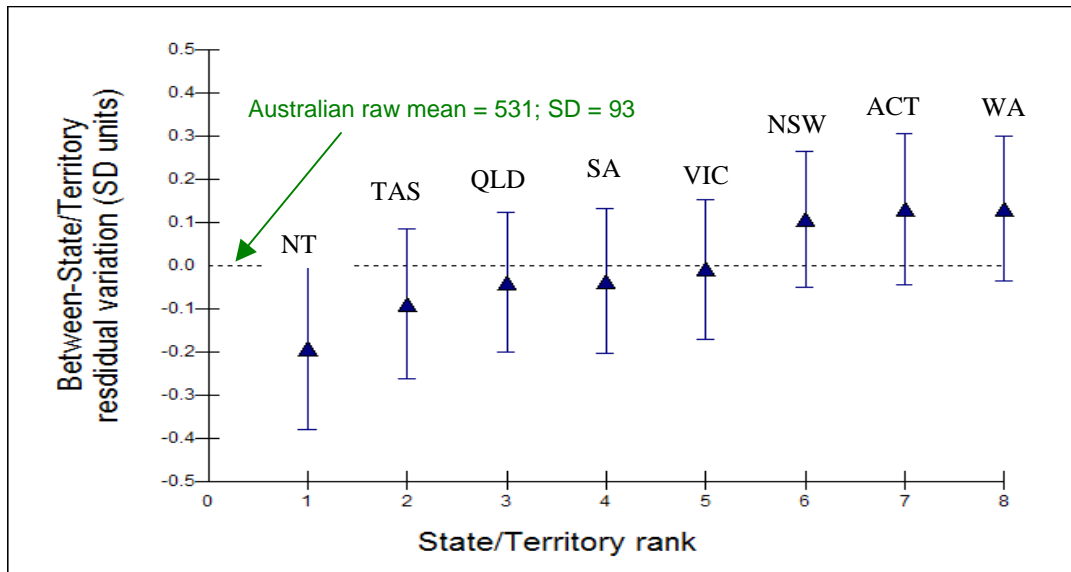


Figure A2.6 Ranked State/Territory-level raw residuals for PISA 2000 *Mathematical Literacy* scores, showing mean-point estimates bounded by 95% ‘uncertainty’ intervals

Comment: With the possible exception of the NT (marginally below), the ‘uncertainty’ intervals around the unadjusted means for other States and Territories all overlap the ‘population’ mean (zero) – indicating non-significant differences at the State/Territory-level in students’ PISA 2000 *Mathematical Literacy* scores – accounting for a mere 1.8% of the residual variance.

In the following model, adjustments are made for the ‘intake’ variables of *Gender*, *Age*, *Grade*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average effect of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses. [Note: statistical significance at the $p < 0.05$ level is indicated when the parameter estimates are at least twice the magnitude of their corresponding standard errors].

$$\text{Mathematics}_{ijk} = \beta_{0ijk}\text{cons} + -0.124(0.034)\text{Gender}_{ijk} + 0.334(0.024)\text{Grade}_{ijk} + 0.213(0.019)\text{SES}_{ijk} + 0.100(0.017)\text{HEDRES}_{ijk} + 0.188(0.024)\text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = 0.070(0.047) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.012(0.008)] \quad \text{Between-State/Territory res. var: 1.6\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.048(0.010)] \quad \text{Between-schools res. var: 6.5\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.680(0.019)] \quad \text{Within-schools res. var: 91.9\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of males), *Grade* (in favour of higher Grade membership), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of Australian students’ PISA 2000 *Mathematical Literacy* achievement scores.

Together, all five fitted variables accounted for 26.2% of the variance in students’ achievement scores, with an insignificant 1.6% of the residual variance at the State/Territory-

level, and a small but significant 6.5% of the residual variance due to variation between schools. As expected, the bulk of the residual variance was at the student-level (i.e., 91.9%). Mean-adjusted residual plots at the State/Territory-level illustrate these results as summarised in Figure A2.7 below.

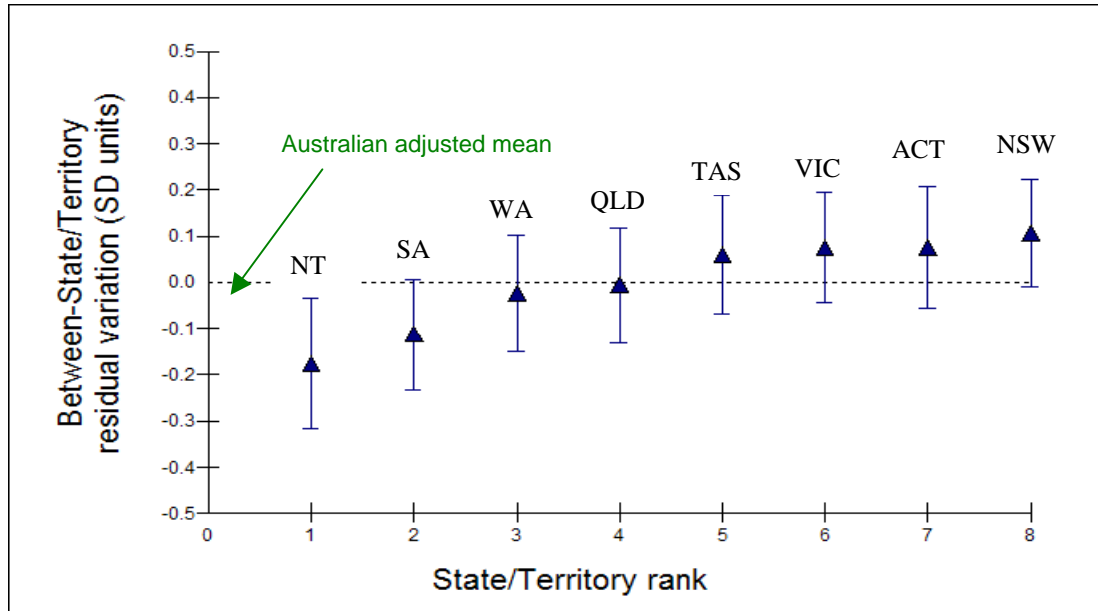


Figure A2.7 Plot of ranked residuals at the State-level, showing adjusted mean-point PISA 2000 *Mathematical Literacy* score estimates, bounded by 95% 'uncertainty' intervals

While these results are of minor interest, they mask the variation between-schools at the national level, as well as between-school variation within each of the States and Territories separately. Following in Figure A2.8 is a plot of ranked residuals for 231 Australian schools from a multilevel analysis of residuals for performance in PISA 2000 *Mathematical Literacy*, showing adjusted mean-point estimates bounded by 95% 'uncertainty' intervals.

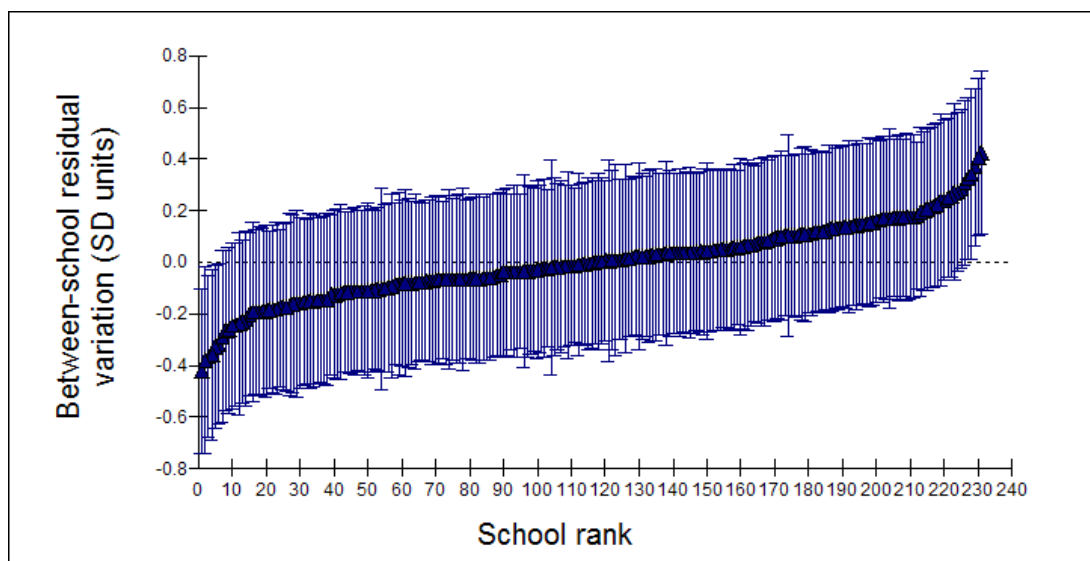


Figure A2.8 Plot of ranked residuals for 231 schools, showing adjusted mean-point PISA 2000 *Mathematical Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: From Figure A2.8, the ‘uncertainty’ intervals around the adjusted means for more than 90% the 231 schools all overlap the national mean (zero) – indicating non-significant differences in students’ achievements between these schools (i.e., a small but significant 6.5% of the residual variance in students’ PISA 2000 *Mathematical Literacy* scores). Nevertheless, the remaining ~10% of schools yielded student performances either significantly above or below the national mean – again indicating differences in the quality of teaching and learning provision among Australian schools.

PISA 2003: Single-level multivariate analyses

The following State/Territory comparisons derive from fitting single-level multivariate models to students’ PISA 2003 *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy* achievement scores (for specific details, see Thomson, Cresswell & De Bortoli, 2004), but adjusted for *State/Territory* location, *Gender*, *Age*, *Grade*, *SES*, and *Home Educational Resources* (HEDRES). As indicated earlier, it is important that these findings from fitting single-level models to the student achievement data are not over-interpreted, since such models ignore the inherent hierarchical structure of the data; i.e., 12,551 students’ achievement scores (level-1) clustered within 321 schools (level-2) and 8 States/Territories (level-3).

PISA 2003 Reading Literacy

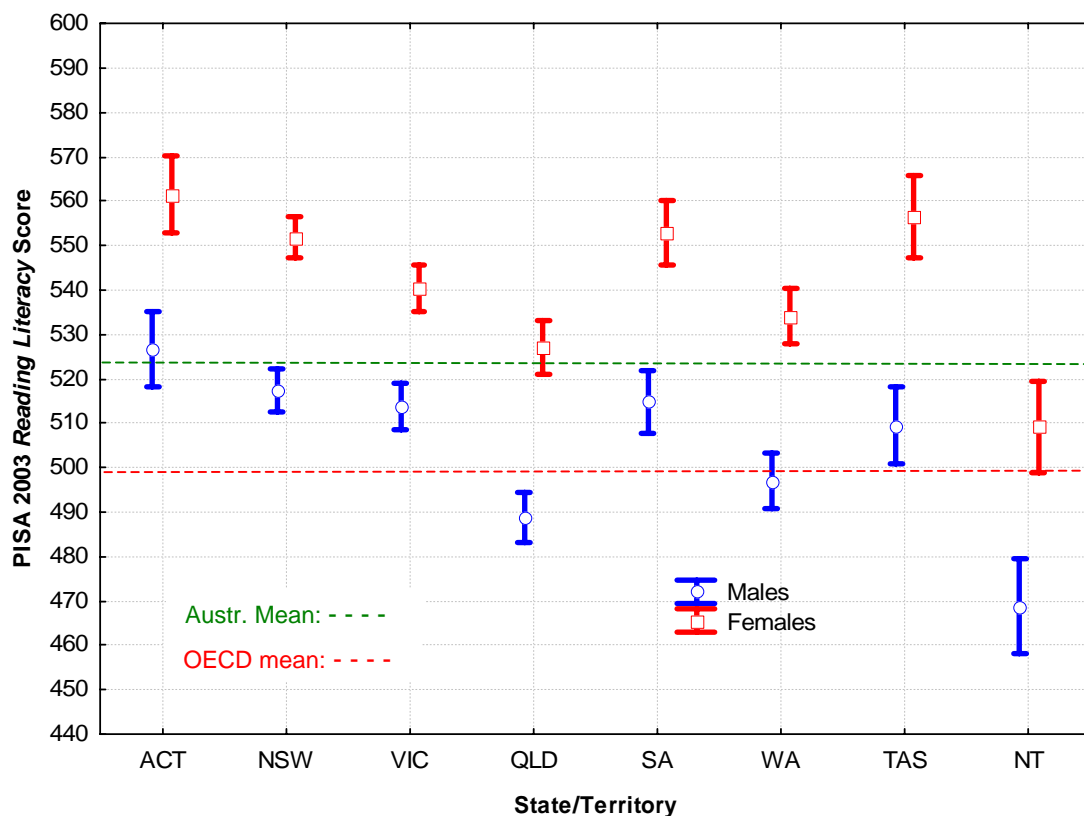


Figure A2.9 Adjusted mean-point estimates of students’ PISA 2003 *Reading Literacy* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,12481) = 33.1, p < 0.000001$
 Gender effect: $F(1,12481) = 394.9, p < 0.000001$
 Age effect: $F(1,12481) = 63.3, p < 0.000001$
 Grade effect: $F(1,12481) = 540.8, p < 0.000001$
 SES effect: $F(1,12481) = 221.9, p < 0.000001$
 HEDRES effect: $F(1,12481) = 763.5, p < 0.000001$
 State/Territory \times Gender effect: $F(7,12481) = 1.5, p = 0.179$ (n.s.)

Comment: With the exception of the *State/Territory* \times *Gender* interaction effect, all main effects were statistically significant, the most notable of which were: *Home Educational Resources* (HEDRES – in favour of higher HEDRES), *Grade* (in favour of higher Grade membership), *Gender* (in favour of females), *SES* (in favour of higher SES), and *Age* (in favour of older students).

Compared with the equivalent findings for PISA 2000 *Reading Literacy* (see Fig. A2.1), the 2003 findings indicated a wider ‘gap’ between the performance levels of females and males (in favour of females) in all States/Territories, including among ACT students where the ‘gender gap’ was not statistically significant in 2000. The means for males in QLD and NT were significantly below the OECD average. This increasing achievement ‘gap’ in favour of females is evident in the raw, unadjusted mean score data summarised in Table A2.1 below

Table A2.1 Female-Male Unadjusted Mean Score Differences for PISA 2000 and 2003 *Reading Literacy*, by Australian States and Territories¹

Mean Female-Male Difference	ACT	NSW	VIC	QLD	SA	WA	TAS	NT
² PISA 2003	42	39	30	49	34	40	45	58
³ PISA 2000	23	30	28	47	29	34	50	30

¹ Adapted from Thomson, Creswell and De Bortoli (2004, p. 105)

² All female-male differences are statistically significant at the $p < 0.05$ level.

³ All female-male differences are statistically significant at the $p < 0.05$ level, except for ACT

PISA 2003 *Mathematical Literacy*

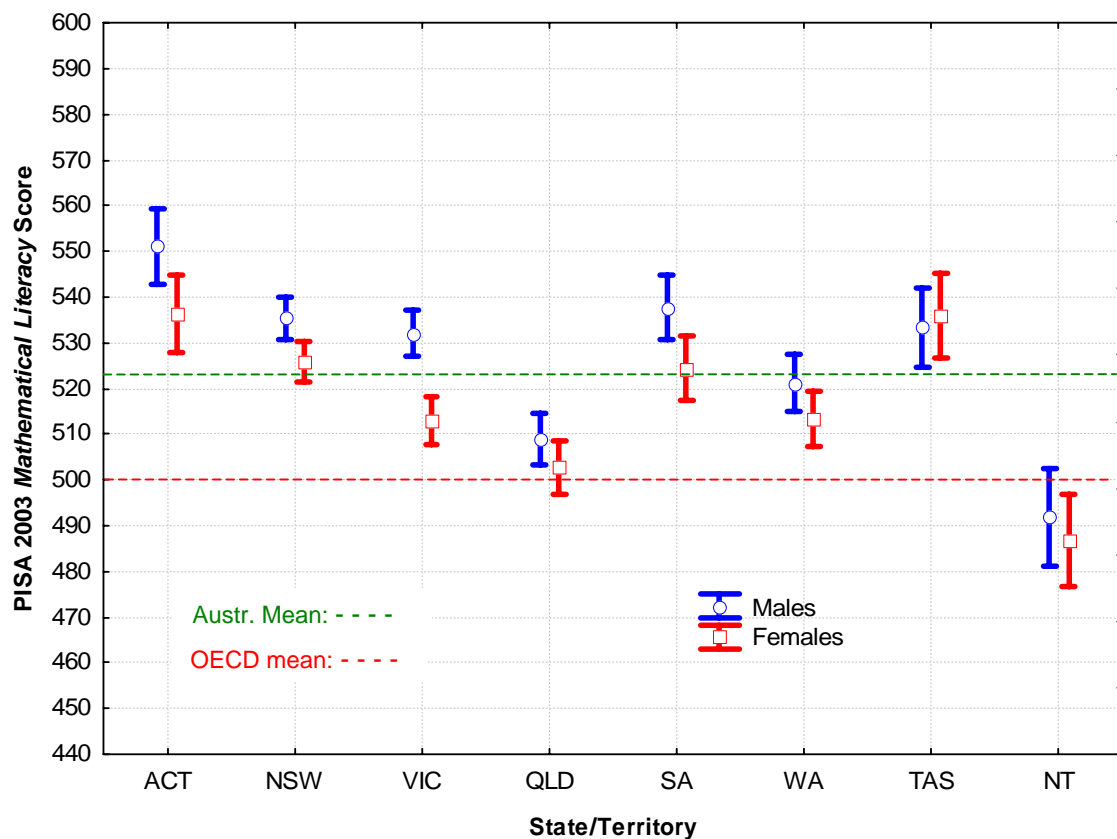


Figure A2.10 Adjusted mean-point estimates of students' PISA 2003 *Mathematical Literacy* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,12481) = 31.7, p < 0.000001$
 Gender effect: $F(1,12481) = 24.9, p < 0.00001$
 Age effect: $F(1,12481) = 54.8, p < 0.000001$
 Grade effect: $F(1,12481) = 650.2, p < 0.000001$
 SES effect: $F(1,12481) = 306.5, p < 0.000001$
 HEDRES effect: $F(1,12481) = 770.1, p < 0.000001$
 State/Territory \times Gender effect: $F(7,12481) = 1.8, p = 0.086$ (n.s.)

Comment: With the exception of the *State/Territory* \times *Gender* interaction effect, all other main effects were statistically significant, the most notable of which were: *Home Educational Resources* (HEDRES – in favour of higher HEDRES), *Grade* (in favour of higher Grade membership), *SES* (in favour of higher *SES*), *State/Territory* membership, *Age* (in favour of older students) and *Gender* (in favour of males).

Apart from NSW and VIC, the gender differences *within* the other States and Territories were not significantly different. Interestingly, the average performance of males in ACT schools was significantly greater than their fellow males in NSW, VIC, QLD, WA and NT. Further, the average performance of females in ACT schools was significantly greater than their female counterparts in VIC, QLD, WA and NT.

PISA 2003 Scientific Literacy

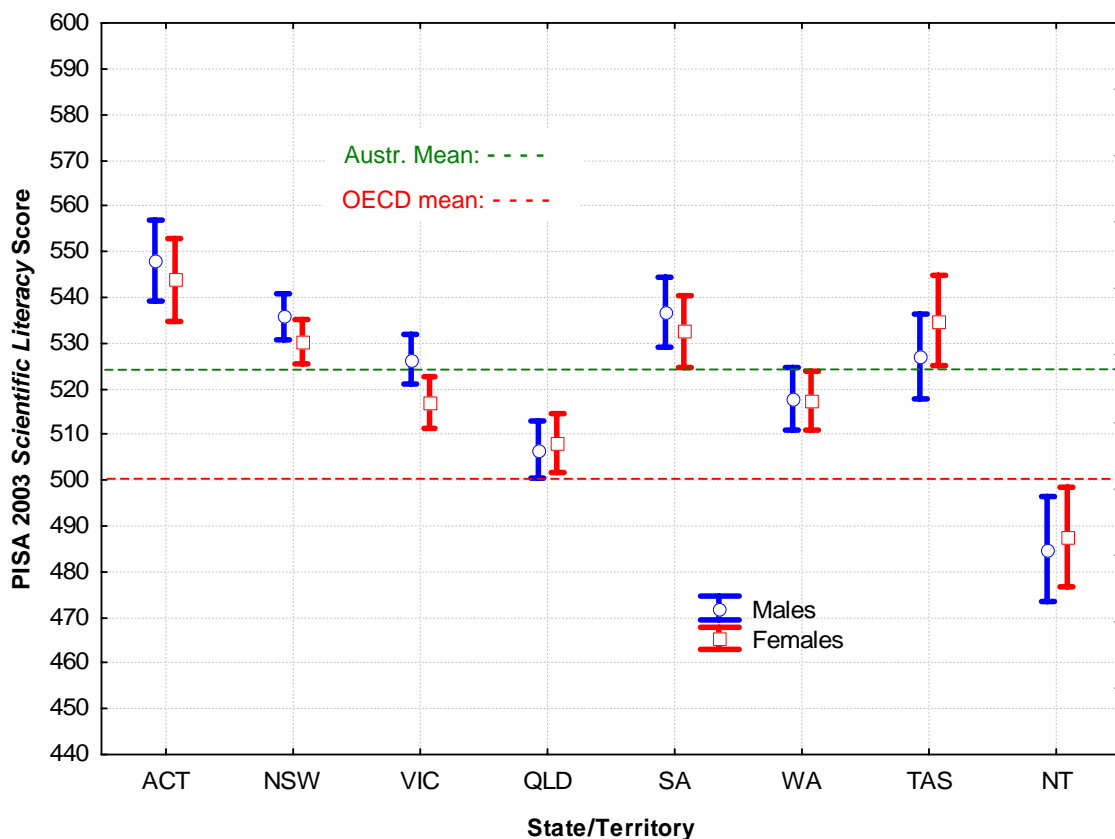


Figure A2.11 Adjusted mean-point estimates of students' *Scientific Literacy* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,12481) = 31.3, p < 0.000001$
 Gender effect: $F(1,12481) = 0.5, p = 0.470$ (n.s.)
 Age effect: $F(1,12481) = 16.3, p = 0.00006$
 Grade effect: $F(1,12481) = 422.4, p < 0.000001$
 SES effect: $F(1,12481) = 283.4, p < 0.000001$
 HEDRES effect: $F(1,12481) = 740.3, p < 0.000001$
 State/Territory \times Gender effect: $F(7,12481) = 1.1, p = 0.377$ (n.s.)

Comment: With the exception of the *State/Territory* \times *Gender* interaction and *Gender* effects, all other main effects were statistically significant, the most notable of which were: *Home Educational Resources* (HEDRES – in favour of higher HEDRES), *Grade* (in favour of higher Grade membership), *SES* (in favour of higher *SES*), *State/Territory*, and *Age* (in favour of older students). Note that the *Gender* differences *within* States and Territories were not significantly different.

The average performance of both males and females in ACT schools was significantly greater than their counterparts in VIC, QLD, WA and NT. Further, the average performance of females in ACT schools was significantly greater than their female counterparts in VIC, QLD, WA and NT. Moreover, the mean performances of both male and female students in ACT, NSW, and SA schools were significantly above the Australian average.

Multilevel analyses of PISA 2003 data

Following are results of the fitted baseline Variance Components model for 12,551 15 year-old students (*i*) in 321 schools (*j*) drawn from 8 Australian States/Territories (*k*) – based on students' normalised scaled scores for PISA 2003 *Reading Literacy*, *Mathematical Literacy* and *Scientific Literacy*. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

$$\text{Reading}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.032(0.056) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.018(0.012) \end{bmatrix} \quad \text{Between-State/Territory res. var: 1.9\% (n.s.)}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.196(0.017) \end{bmatrix} \quad \text{Between-schools res. var: 20.8\%}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.792(0.010) \end{bmatrix} \quad \text{Within-schools res. var: 77.3\%}$$

Comment: Around the normalised grand mean for Australia (-0.032), the residual variance in students' PISA 2003 *Reading Literacy* achievement scores between States and Territories was **not** statistically significant (with the possible exception of NT). The unadjusted residual plots at the State/Territory-level illustrate these results as follows:

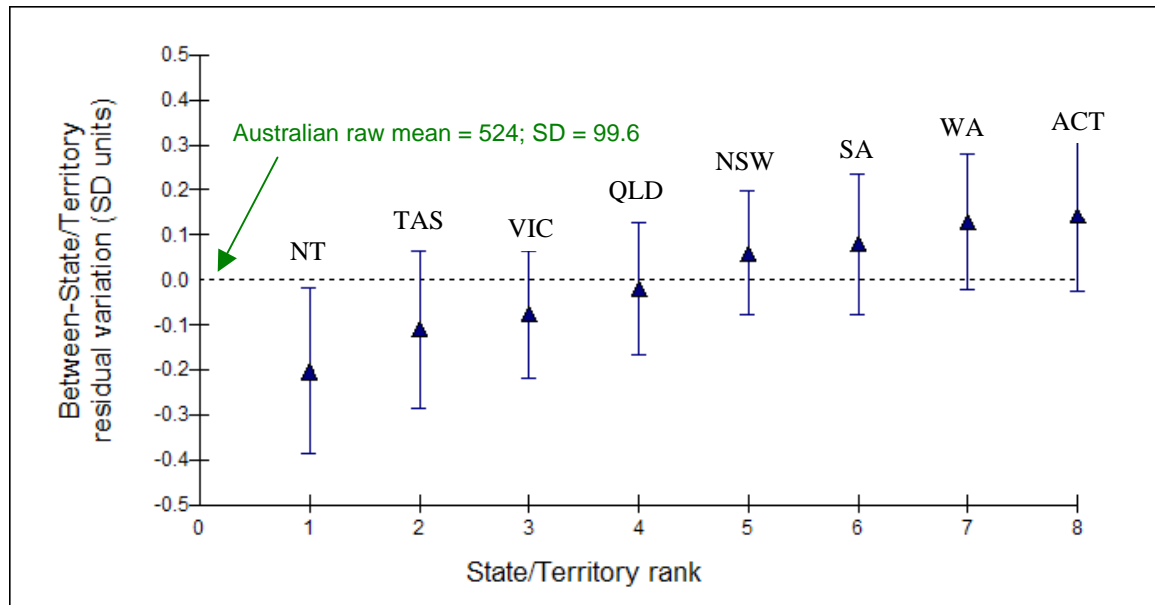


Figure A2.12 Ranked State/Territory-level raw residuals for PISA 2003 *Reading Literacy* scores, showing mean-point estimates, bounded by 95% ‘uncertainty’ intervals

Comment: Apart from NT (below), the ‘uncertainty’ intervals around the unadjusted means for other States and Territories all overlap the ‘population’ mean (zero) – indicating non-significant differences at the State/Territory-level in students’ PISA 2003 *Reading Literacy* scores – accounting for a mere 1.9% of the residual variance. Whereas the estimates obtained from fitting a variance-components model to the data are not of particular interest (*per se*), they provide a useful baseline from which to compare findings from fitting more ‘responsible’ models.

In the following model, adjustments are made for the ‘intake’ variables of *Gender*, *Age*, *Grade*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average ‘cohort effect’ of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses. [Note: statistical significance at the $p < 0.05$ level is indicated when parameter estimates are at least twice the magnitude of their corresponding standard errors].

$$\text{Reading}_{ijk} = \beta_{0ijk} \text{Cons} + 0.352(0.017) \text{Gender}_{ijk} + -0.069(0.009) \text{AGE}_{ijk} + 0.310(0.013) \text{Grade}_{ijk} + 0.076(0.008) \text{SES}_{ijk} + 0.192(0.010) \text{HEDRES}_{ijk} + 0.208(0.019) \text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = -0.194(0.057) + v_{0jk} + u_{0jk} + e_{0ijk}$$

$$[v_{0jk}] \sim N(0, \Omega_v) : \Omega_v = [0.022(0.013)] \quad \text{Between-State/Territory res. var: 2.7\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.094(0.009)] \quad \text{Between-schools res. var: 11.5\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.703(0.009)] \quad \text{Within-schools res. var: 85.8\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of females), *Grade* (in favour of higher Grade membership), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of

Australian students' PISA 2003 *Reading Literacy* achievement scores. While the effect of *Age* was small, it was statistically significant (in favour of younger students).

Together, all six fitted variables accounted for only 13.2% of the variance in students' achievement scores, with an insignificant 2.7% of the residual variance at the State/Territory-level, and a significant 11.5% of the residual variance due to variation between schools. As expected, the bulk of the residual variance was at the student-level (i.e., 85.8%). Mean-adjusted residual plots at the State/Territory-level illustrate these results are presented in Figure A2.13 below.

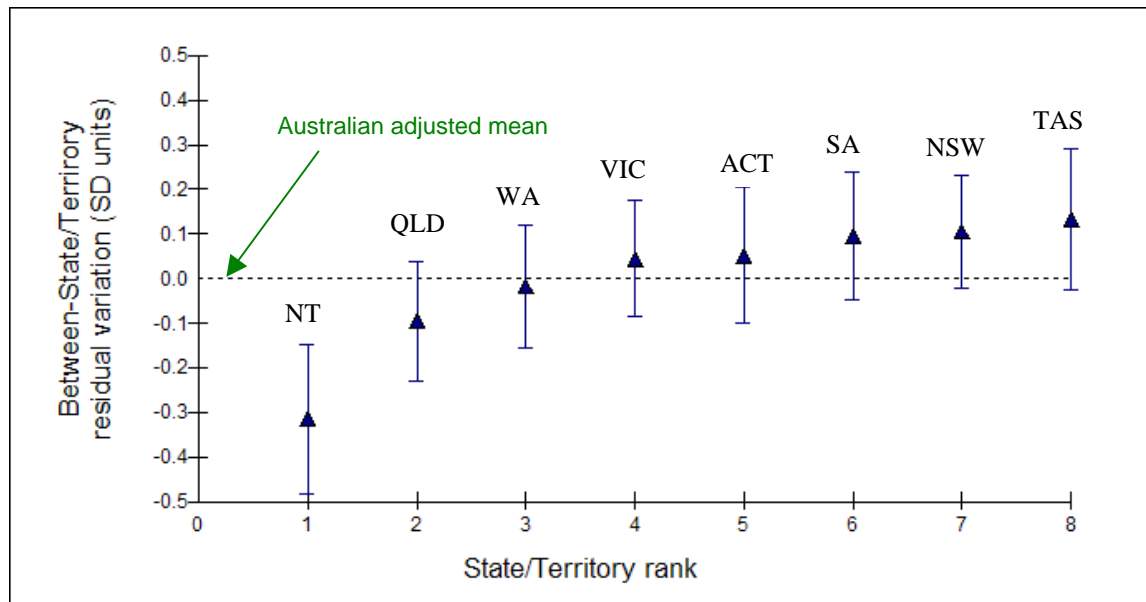


Figure A2.13 Plot of ranked State/Territory residuals, showing adjusted mean-point PISA 2003 *Reading Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: With the exception of NT (below), the 95% 'uncertainty' intervals around the adjusted means, the intervals for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., a small 2.7% of the residual variance in students' PISA 2003 *Reading Literacy* scores). Indeed, there were no significant differences between the adjusted mean performances of students located in QLD, WA, VIC, ACT, SA, NSW, and TAS schools.

Although these results are of minor interest, they mask the variation between-schools at the national level, as well as between-school variation within each of the States and Territories separately. Following in Figure A2.14 is a plot of ranked residuals for 321 Australian schools from a multilevel analysis of residuals for performance in PISA 2003 *Reading Literacy*, showing adjusted mean-point estimates bounded by 95% 'uncertainty' intervals.

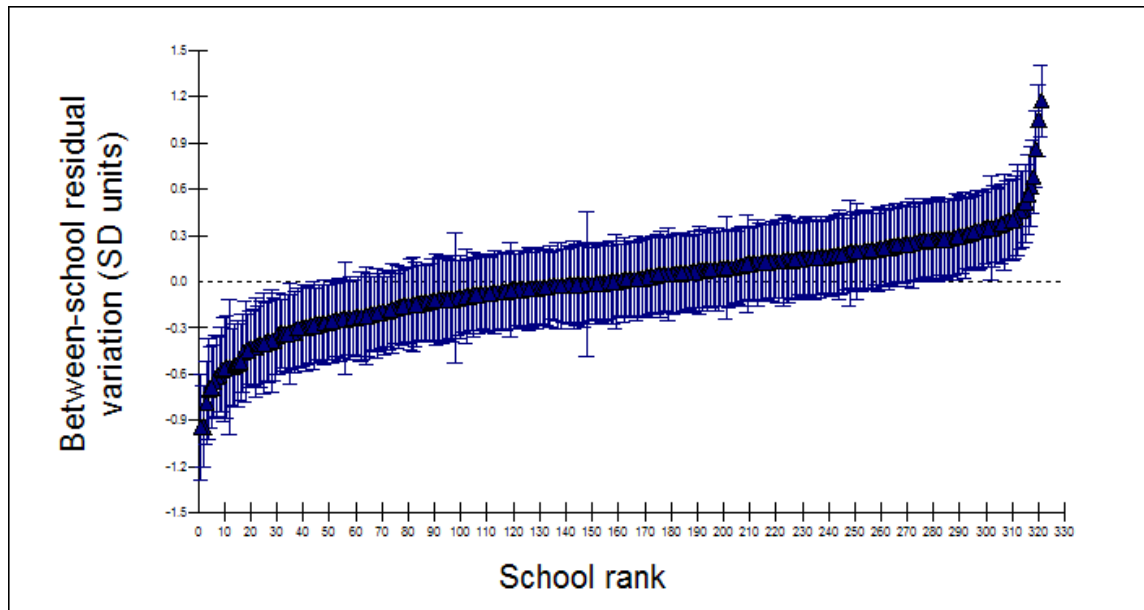


Figure A2.14 Plot of ranked residuals for 321 schools, showing adjusted mean-point PISA 2003 *Reading Literacy* score estimates, bounded by 95% ‘uncertainty’ intervals

Comment: From Figure A2.14, the ‘uncertainty’ intervals around the adjusted means for approximately 85% the 321 schools all overlap the national mean (zero) – indicating non-significant differences in students’ *Reading Literacy* scores between these schools. The remaining 15% of schools yielded student performances either significantly above or below the national mean, resulting in an increased between-school residual variation (11.5%) compared with PISA 2000 *Reading Literacy* (5.5%; see Fig.A2.5). These findings again indicated differences in the quality of teaching and learning provision among Australian schools.

PISA 2003 *Mathematical Literacy*

Following are results of the fitted baseline Variance Components model for 12,551 15 year-old students (i) in 321 schools (j) drawn from 8 Australian States/Territories (k) – based on students’ normalised scaled scores for PISA 2003 *Mathematical Literacy*. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

$$\text{Mathematics}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.024(0.060) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.022(0.014) \end{bmatrix} \text{ Between-State/Territory res. var: 2.2\% (n.s.)}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.199(0.018) \end{bmatrix} \text{ Between-schools res. var: 19.8\%}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.785(0.010) \end{bmatrix} \text{ Within-schools res. var: 78.0\%}$$

Comment: Around the normalised grand mean (-0.024), the residual variance in students’ PISA 2003 *Mathematical Literacy* achievement scores between States and Territories were **not**

statistically significant. Unadjusted residual plots at the State/Territory-level illustrate these results as presented in Figure A2.15 below.

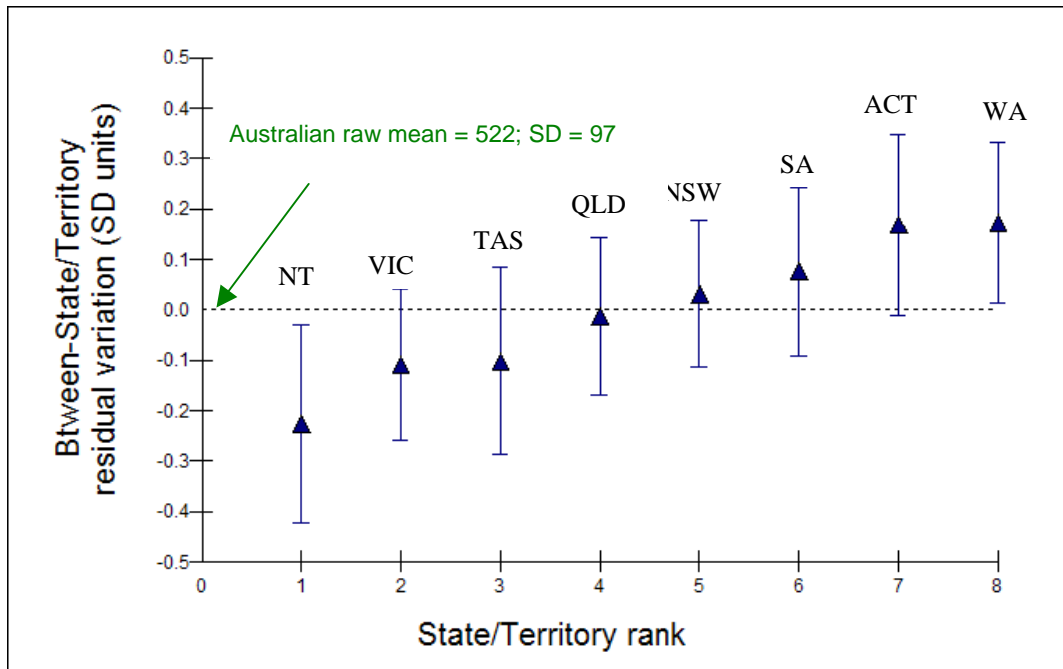


Figure A2.15 Ranked State/Territory-level raw residuals for PISA 2003 *Mathematical Literacy* scores, showing mean-point estimates bounded by 95% 'uncertainty' intervals

Comment: Apart from NT (below) and WA (above), the 'uncertainty' intervals around the unadjusted means for other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level in students' PISA 2003 *Mathematical Literacy* scores – accounting for a small 2.2% of the residual variance.

In the following model, adjustments are made for the 'intake' variables of *Gender*, *Age*, *Grade*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average effect of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses. [Note: statistical significance at the $p < 0.05$ level is indicated when the parameter estimates are at least twice the magnitude of their corresponding standard errors].

$$\text{Mathematics}_{ijk} = \beta_{0ijk} \text{Cons} + -0.107(0.017) \text{Gender}_{ijk} + -0.068(0.009) \text{AGE}_{ijk} + 0.350(0.014) \text{Grade}_{ijk} + 0.088(0.008) \text{SES}_{ijk} + 0.200(0.010) \text{HEDRES}_{ijk} + 0.216(0.019) \text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = 0.042(0.058) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.023(0.013)] \quad \text{Between-State/Territory res. var: 2.8\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.096(0.009)] \quad \text{Between-schools res. var: 11.6\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.709(0.009)] \quad \text{Within-schools res. var: 85.6\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of males), *Age* (in favour of younger students), *Grade* (in favour of higher Grade membership), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of Australian students' PISA 2003 *Mathematical Literacy* achievement scores.

Together, all six fitted variables accounted for 17.7% of the variance in students' achievement scores, with an insignificant 2.8% of the residual variance at the State/Territory-level, but a significant 11.6% of the residual variance due to variation between schools. As expected, the bulk of the residual variance was at the student-level (i.e., 85.6%). Mean-adjusted residual plots at the State/Territory-level illustrate these results as follows.

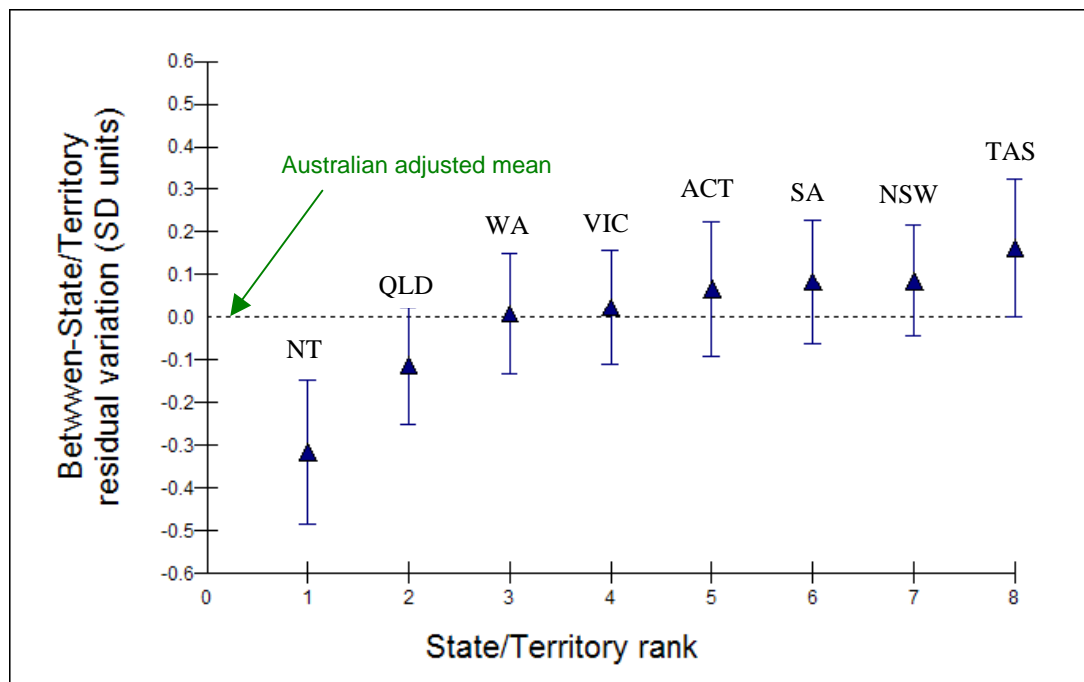


Figure A2.16 Plot of ranked State/Territory residuals, showing adjusted mean-point PISA 2003 *Mathematical Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: With the exception of the NT (below) and TAS (marginally above), the 95% 'uncertainty' intervals around the adjusted means for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., an insignificant 2.8% of the residual variance in students' PISA 2000 *Mathematical Literacy* scores). Indeed, there were no significant differences between the adjusted mean performances of students located in QLD, WA, VIC, ACT, SA, NSW and TAS schools.

Although these results are of minor interest, they mask the variation between-schools at the State/Territory-level, as well as within each of the States and Territories separately. Following in Figure A2.17 is a plot of ranked residuals for 321 Australian schools from a multilevel analysis of residuals for performance in PISA 2003 *Mathematical Literacy*, showing adjusted mean-point estimates bounded by 95% 'uncertainty' intervals.

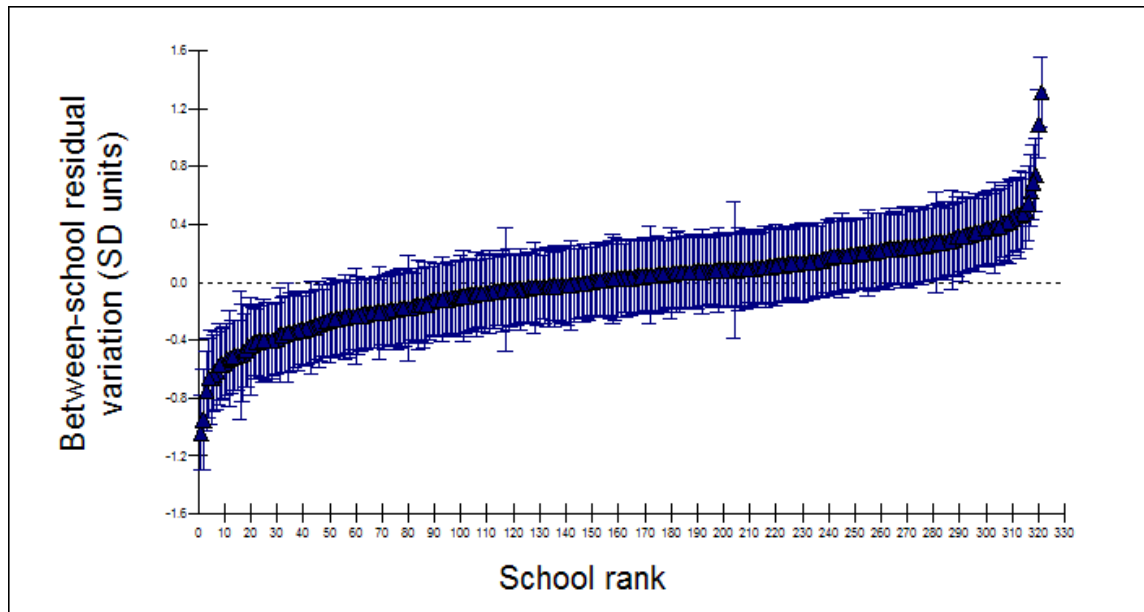


Figure A2.17 Plot of ranked residuals for 321 schools, showing adjusted mean-point PISA 2003 *Mathematical Literacy* score estimates, bounded by 95% ‘uncertainty’ intervals

Comment: From Figure A2.17, the ‘uncertainty’ intervals around the adjusted means for more than 80% the 321 schools all overlap the national mean (zero) – indicating non-significant differences in students’ achievements between these schools. The remaining ~20% of schools yielded student performances either significantly above or below the national mean, resulting in an increased between-school residual variation (11.6%) compared with PISA 2000 *Mathematical Literacy* (6.5%; see Fig.A2.8). These findings again indicated differences in the quality of teaching and learning provision among Australian schools.

PISA 2003 *Scientific Literacy*

Following are results of the fitted baseline Variance Components model for 12,551 15 year-old students (i) in 321 schools (j) drawn from 8 Australian States/Territories (k) – based on students’ normalised scaled scores for PISA 2003 *Scientific Literacy*. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

$$\text{Science}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.028(0.060) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.023(0.014) \end{bmatrix} \text{ Between-State/Territory res. var: 2.3\% (n.s.)}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.183(0.016) \end{bmatrix} \text{ Between-schools res. var: 18.1\%}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.803(0.010) \end{bmatrix} \text{ Within-schools res. var: 79.6\%}$$

Comment: Around the normalised grand mean (-0.024), the residual variance in students’ PISA 2003 *Scientific Literacy* achievement scores between States and Territories were **not** statistically significant. Unadjusted residual plots at the State/Territory-level illustrate these results as follows:

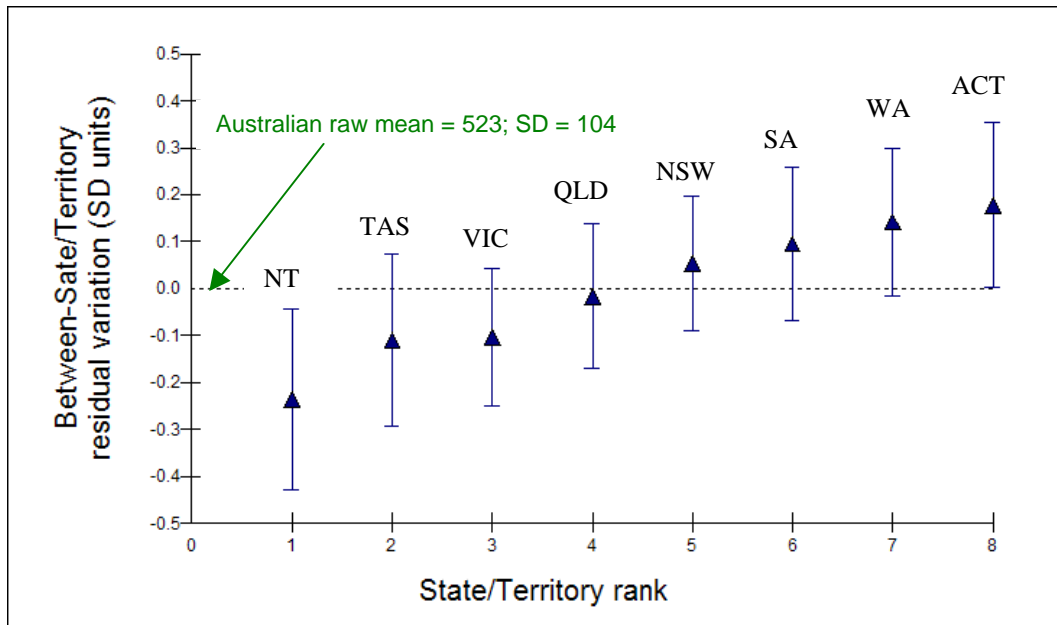


Figure A2.18 Ranked State/Territory raw residuals for PISA 2003 *Scientific Literacy* scores, showing mean-point estimates bounded by 95% ‘uncertainty’ intervals

Comment: With the exception of the NT (below) and the ACT (marginally above), the ‘uncertainty’ intervals around the unadjusted means for other States all overlap the ‘population’ mean (zero) – indicating non-significant differences at the State/Territory-level in students’ PISA 2003 *Mathematical Literacy* scores – accounting for a non-significant 2.3% of the residual variance.

In the following model, adjustments are made for the ‘intake’ variables of *Gender*, *Age*, *Grade*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average effect of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses.

$$\text{Science}_{ijk} = \beta_{0ijk} \text{Cons} + -0.032(0.017) \text{Gender}_{ijk} + -0.035(0.010) \text{AGE}_{ijk} + 0.286(0.014) \text{Grade}_{ijk} + 0.087(0.009) \text{SES}_{ijk} + 0.204(0.010) \text{HEDRES}_{ijk} + 0.210(0.019) \text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = 0.002(0.057) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.022(0.013)] \quad \text{Between-State/Territory res. var: 2.6\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.088(0.009)] \quad \text{Between-schools res. var: 10.4\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.738(0.009)] \quad \text{Within-schools res. var: 87.0\%}$$

Comment: These results indicate significant effects for: *Age* (in favour of younger students), *Grade* (in favour of higher Grade membership), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of Australian students’ PISA 2003 *Scientific Literacy* achievement scores. Note that the *Gender* effect (marginally in favour of males) was not statistically significant.

Together, the six fitted variables accounted for 16% of the variance in students' achievement scores, with an insignificant 2.8% of the residual variance at the State/Territory-level, and a significant 10.4% of the residual variance due to variation between schools. As expected, the bulk of the residual variance was at the student-level (i.e., 87.0%). Mean-adjusted residual plots at the State/Territory-level illustrate these findings in Figure A2.19 below.

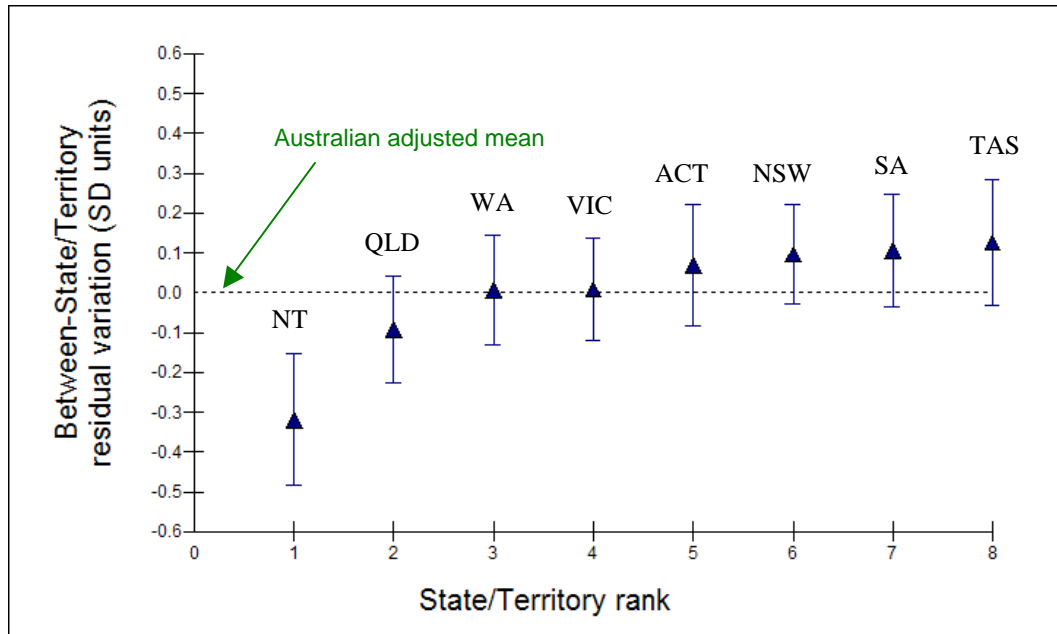


Figure A2.19 Plot of ranked State/Territory residuals, showing adjusted mean-point PISA 2003 *Scientific Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: From Figure A2.19, with the exception of the NT (below), the 95% 'uncertainty' intervals around the adjusted means for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., an insignificant 2.8% of the residual variance in students' PISA 2003 *Scientific Literacy* scores). Indeed, there were no significant differences between the adjusted mean performances of students located in QLD, WA, VIC, ACT, SA, NSW and TAS schools.

Whereas these results are of minor interest, they mask the variation between-schools at the State/Territory-level, as well as within each of the States and Territories separately. Following in Figure A2.20 is a plot of ranked residuals for 321 Australian schools from a multilevel analysis of residuals for performance in PISA 2003 *Scientific Literacy*, showing adjusted mean-point estimates bounded by 95% 'uncertainty' intervals.

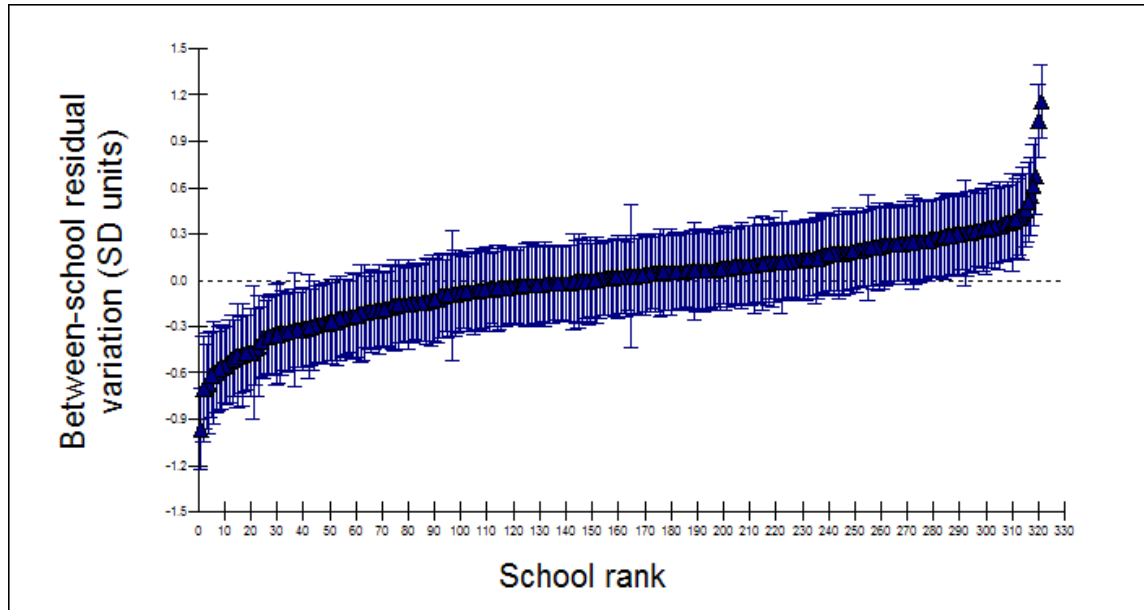


Figure A2.20 Plot of ranked residuals for 321 schools, showing adjusted mean-point PISA 2003 *Mathematical Literacy* score estimates, bounded by 95% 'uncertainty' intervals

Comment: From Figure A2.20, the 'uncertainty' intervals around the adjusted means for more than 80% the 321 schools all overlap the national mean (zero) – indicating non-significant between-school differences in students' PISA 2003 *Scientific Literacy* scores. The remaining ~20% of schools yielded student performances either significantly above or below the national mean, resulting in a significant between-school residual variation (10.4%). These findings again indicated differences in the quality of teaching and learning provision among Australian schools.

Appendix A3

State/Territory Comparisons of Students' Achievements in TIMSS

TIMSS 2003: Single-level multivariate analyses

The following State/Territory comparisons derive from fitting single-level multivariate models to the *Third International Mathematics and Science Study* (TIMSS 2003) **Mathematics** and **Science** achievement score data. These data were obtained from a designed sample of students in their ninth year of formal schooling within 207 Australian government, Catholic and independent schools (for specific details, see: Martin *et al.*, 2004; Mullis *et al.*, 2004).¹⁴ For present purposes, the achievement scores are adjusted for *State/Territory* location, *Gender*, *Age*, *Grade*, *SES*, and *Home Educational Resources* (HEDRES).

TIMSS 2003 Mathematics

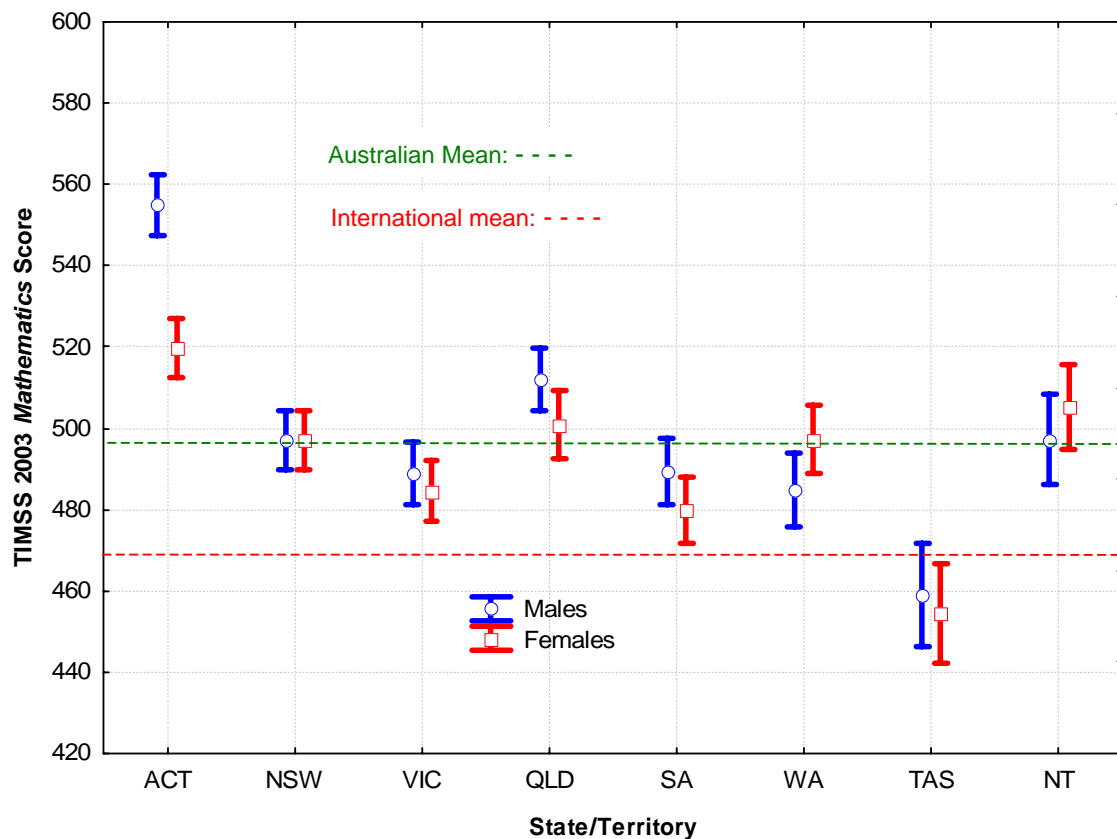


Figure A3.1 Adjusted mean-point estimates for TIMSS 2003 *Mathematics* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,4622) = 48.8, p < 0.000001$

Gender effect: $F(1,4622) = 6.1, p < 0.014$

Age effect: $F(1,4622) = 13.5, p < 0.001$

SES effect: $F(1,4622) = 0.4, p = 0.514$ (n.s.)

HEDRES effect: $F(1,4622) = 268.4, p < 0.000001$

State/Territory \times Gender effect: $F(7,4622) = 6.5, p < 0.000001$

Comment: Whereas the plotted differences are misleading because their standard errors are based on the total number of participating Australian students rather than within each of the 8

¹⁴ For specific details of Australian students' mathematics and science achievements in TIMSS (during their ninth year of formal schooling) for 1998, see Zammit, Routitsky and Greenwood (2002).

States and Territories, the findings from fitting this single-level multivariate model to the TIMSS 2003 *Mathematics* achievement score data indicate significant differences between the States and Territories.

However, of particular note is that on average, both male and female students in ACT schools and males in QLD schools achieved significantly above the Australian and International means. Moreover, with the exception of students in TAS schools, students in all other States and Territories achieved significantly higher scores than the International mean. Further, with the exception of *SES* at the student level, all other effects of the fitted variables were statistically significant, the most notable of which were: *Home Educational Resources* (HEDRES – in favour of higher HEDRES), *Age* (in favour of older students) and *Gender* (in favour of males).

TIMSS 2003 *Science* (Australia)

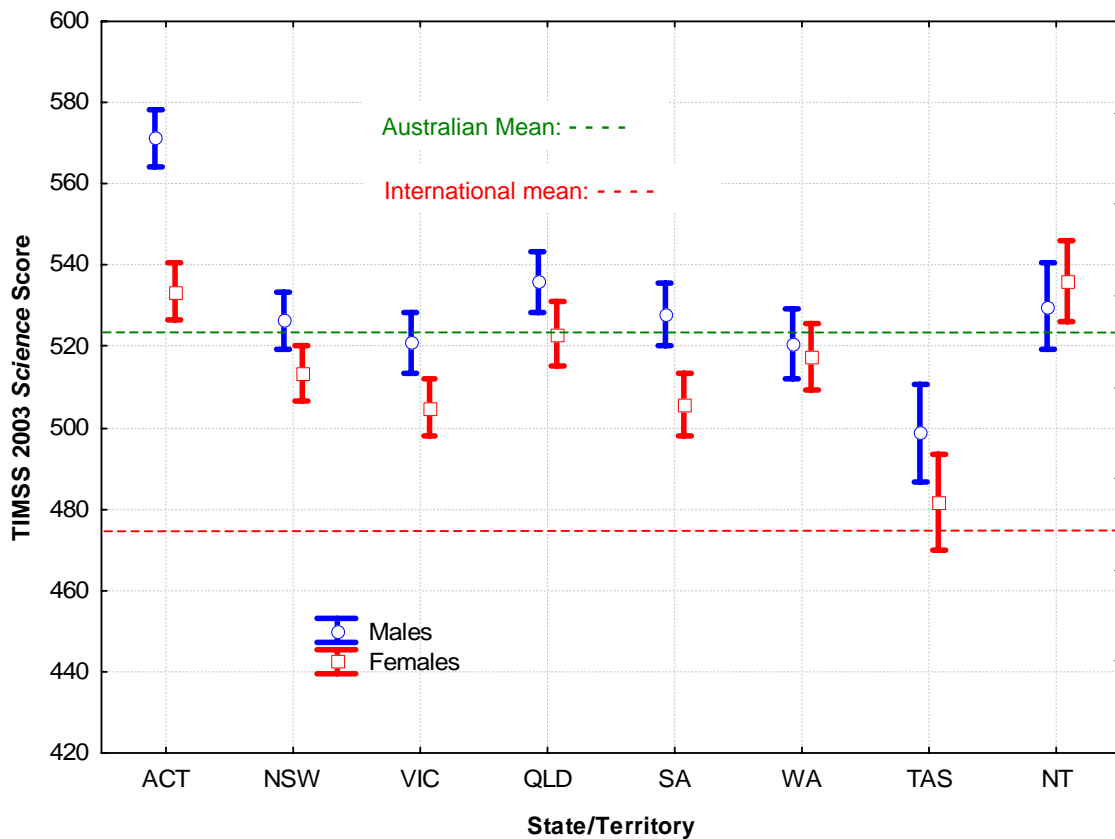


Figure A3.2 Adjusted mean-point estimates for TIMSS 2003 *Science* scores bounded by 95% confidence intervals, by State/Territory and Gender

State/Territory effect: $F(7,4622) = 31.1, p < 0.000001$
 Gender effect: $F(1,4622) = 44.8, p < 0.00001$
 Age effect: $F(1,4622) = 20.3, p < 0.00001$
 SES effect: $F(1,4622) = 2.0, p = 0.159$ (n.s.)
 HEDRES effect: $F(1,4622) = 306.6, p < 0.000001$
 State/Territory \times Gender effect: $F(7,4622) = 4.9, p < 0.0001$

Comment: The findings from fitting this single-level multivariate model to the TIMSS 2003 *Science* achievement data indicate significant differences between the States and Territories. As for the TIMSS 2003 *Mathematics* achievement data (see Fig. A3.1), on average, both male and female students in ACT schools, males in QLD schools and females in NT schools achieved significantly above the Australian and International means. Moreover, with the exception of female students in TAS schools, students in all other States and Territories achieved significantly higher scores than the International mean. Further, with the exception of *SES* at the student

level, all other effects of the fitted variables were statistically significant, the most notable of which were: *Home Educational Resources* (HEDRES – in favour of higher HEDRES), *Age* (in favour of older students) and *Gender* (in favour of males).

Multilevel analyses of TIMSS 2003 data

Following are results of the fitted baseline Variance Components model for 4791 students (*i*) in their ninth year of formal schooling in 207 schools (*j*) drawn from 8 Australian States/Territories (*k*) – based on students' normalised performance scores for TIMSS 2003 *Mathematics* and *Science*. The results present the normalised parameter estimates (coloured green) and their standard errors in parentheses (also coloured green) for the **residual** variation (res. var.) at: the State/Territory-level, between-school-level, and within-school (student)-level.

TIMSS 2003 *Mathematics* (Australia)

$$\text{Mathematics}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.083(0.078) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.028(0.024) \end{bmatrix} \text{ Between-State/Territory res. var: 2.8\% (n.s.)}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.461(0.049) \end{bmatrix} \text{ Between-schools res. var: 45.7\%}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.520(0.011) \end{bmatrix} \text{ Within-schools res. var: 55.5\%}$$

Comment: Around the normalised grand mean (-0.083), the residual variance in students' TIMSS 2003 *Mathematics* achievement scores between States and Territories were **not** statistically significant. Unadjusted residual plots at the State/Territory-level follows.

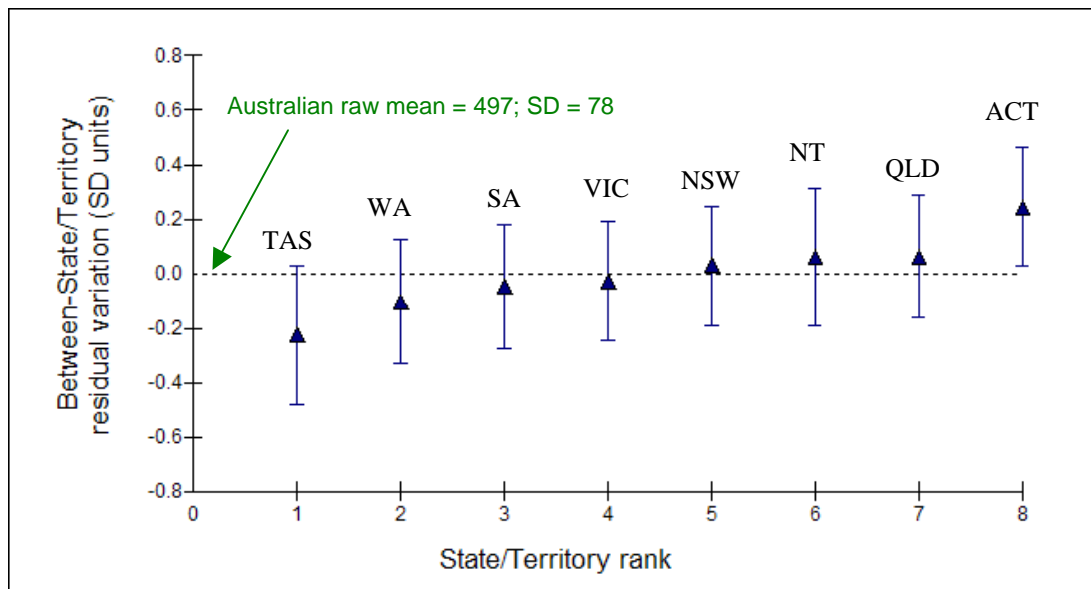


Figure A3.3 Ranked State/Territory raw residuals for TIMSS 2003 *Mathematics* scores, showing mean-point estimates bounded by 95% 'uncertainty' intervals

Comment: With the exception of the ACT (above), the 'uncertainty' intervals around the unadjusted means for other States all overlap the 'population' mean (zero) – indicating non-

significant differences at the State/Territory-level in students' TIMSS 2003 *Mathematics* scores – accounting for a non-significant 2.8% of the residual variance.

In the following model, adjustments are made for the 'intake' variables of *Gender*, *Age*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average effect of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses.

$$\text{Mathematics}_{ijk} = \beta_{0ijk} \text{Cons} + -0.051(0.023) \text{Gender}_{ijk} + -0.047(0.014) \text{Age}_{ijk} + 0.053(0.017) \text{SES}_{ijk} + 0.122(0.019) \text{HEDRES}_{ijk} + 0.229(0.043) \text{ScAvSES}_{jk}$$

$$\beta_{0ijk} = -0.043(0.081) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.037(0.026)] \quad \text{Between-State/Territory res. var: 4.2\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.342(0.037)] \quad \text{Between-schools res. var: 38.7\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.504(0.011)] \quad \text{Within-schools res. var: 57.1\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of males), *Age* (in favour of younger students), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of Australian students' TIMSS 2003 *Mathematics* achievement scores.

Together, the five fitted variables accounted for 12.5% of the variance in students' *Mathematics* achievement scores, with an insignificant 4.2% of the residual variance at the State/Territory-level, a large significant 38.7% of the residual variance due to variation between schools, and 57.1% at the student-level. In instances where there are large residual variances at the school-level (as in the present case), differential quality in teaching and learning provision between schools is typically indicated (see Hill & Rowe, 1996, 1998). Mean-adjusted residual plots at the State/Territory-level illustrate these findings in Figure A3.4 below.

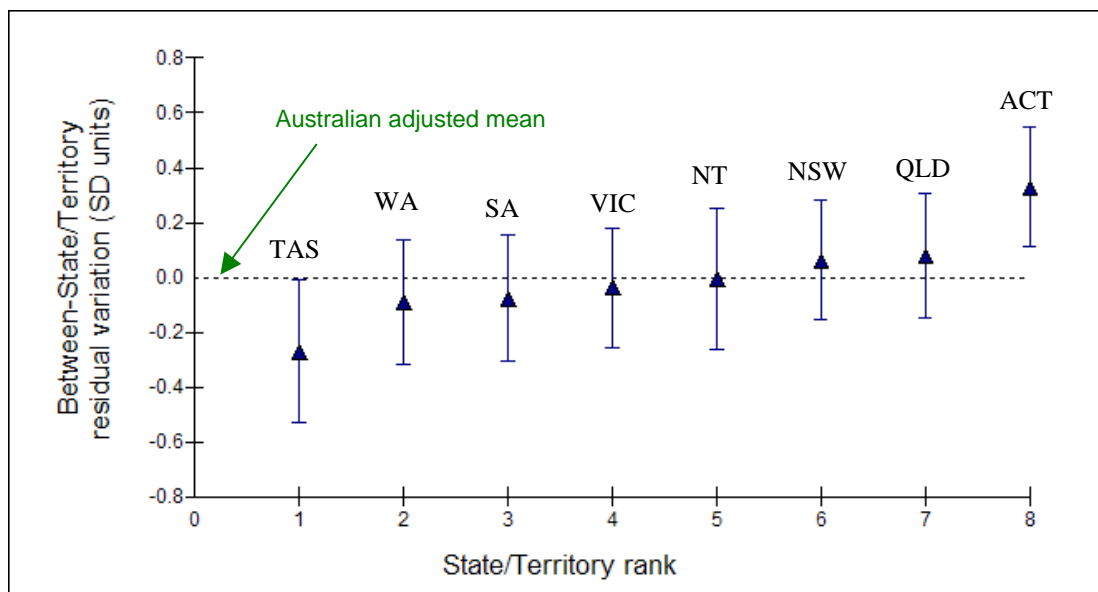


Figure A3.4 Plot of ranked State/Territory residuals, showing adjusted mean-point TIMSS 2003 *Mathematics* score estimates, bounded by 95% 'uncertainty' intervals

Comment: From Figure A3.4, with the exception of TAS (marginally below) and ACT (above), the 95% ‘uncertainty’ intervals around the adjusted means for the other States and Territories all overlap the ‘population’ mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., an insignificant 4.2% of the residual variance in students’ TIMSS 2003 *Mathematics* scores). While on average, students in ACT school performed significantly above the adjusted national mean and better than TAS students, there were no significant differences between the adjusted mean performances of students located in WA, SA, VIC, NT, NSW, QLD and ACT schools.

Whereas these results are of minor interest, they mask the variation between-schools at the State/Territory-level, as well as within each of the States and Territories separately. Following in Figure A3.5 is a plot of ranked residuals for 207 Australian schools from a multilevel analysis of residuals for performance in TIMSS 2003 *Mathematics*, showing adjusted mean-point estimates bounded by 95% ‘uncertainty’ intervals.

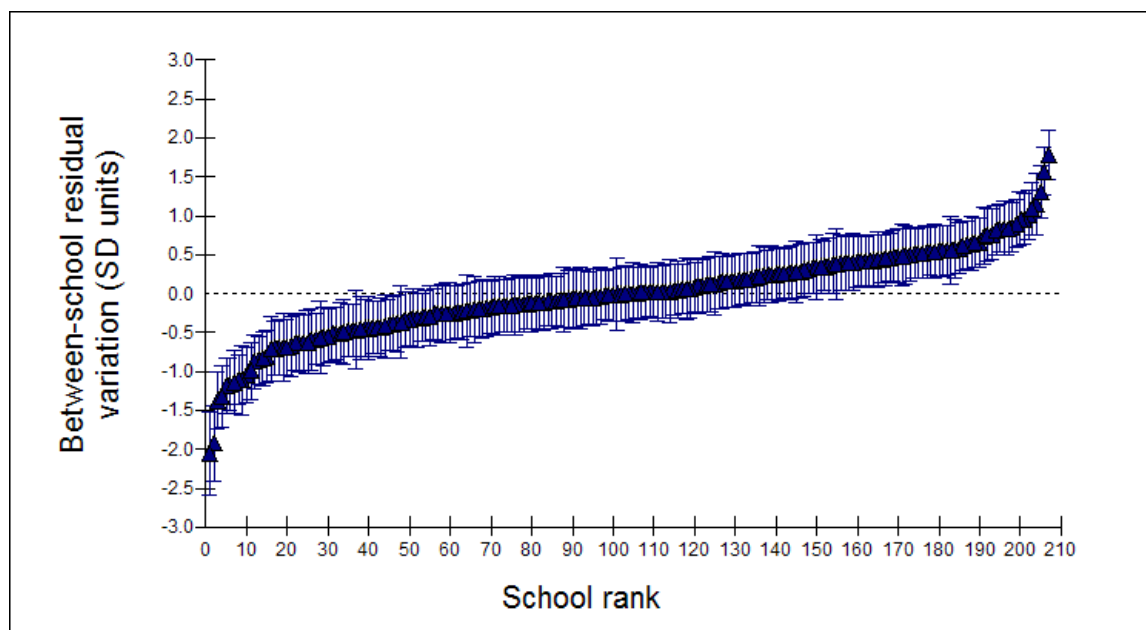


Figure A3.5 Plot of ranked residuals for 207 schools, showing adjusted mean-point TIMSS 2003 *Mathematics* score estimates, bounded by 95% ‘uncertainty’ intervals

Comment: From Figure A3.5, the ‘uncertainty’ intervals around the adjusted means for approximately 55% the 207 schools all overlap the national mean (zero) – indicating non-significant between-school differences in students’ TIMSS 2003 *Mathematics* scores for these schools. The remaining ~45% of schools yielded student performances either significantly above or below the national mean, resulting in a large and significant between-school residual variation (38.7%). Again, these findings indicated differences in the quality of teaching and learning provision among Australian schools.

TIMSS 2003 Science

$$\text{Science}_{ijk} = \beta_{0ijk} \text{Cons}$$

$$\beta_{0ijk} = -0.061(0.069) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.023(0.019)] \quad \text{Between-State/Territory res. var: 2.3\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.349(0.038)] \quad \text{Between-schools res. var: 34.9\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.628(0.013)] \quad \text{Within-schools res. var: 62.8\%}$$

Comment: Around the normalised grand mean (-0.061), the residual variance in students' TIMSS 2003 *Science* achievement scores between States and Territories were **not** statistically significant. Unadjusted residual plots at the State/Territory-level follows.

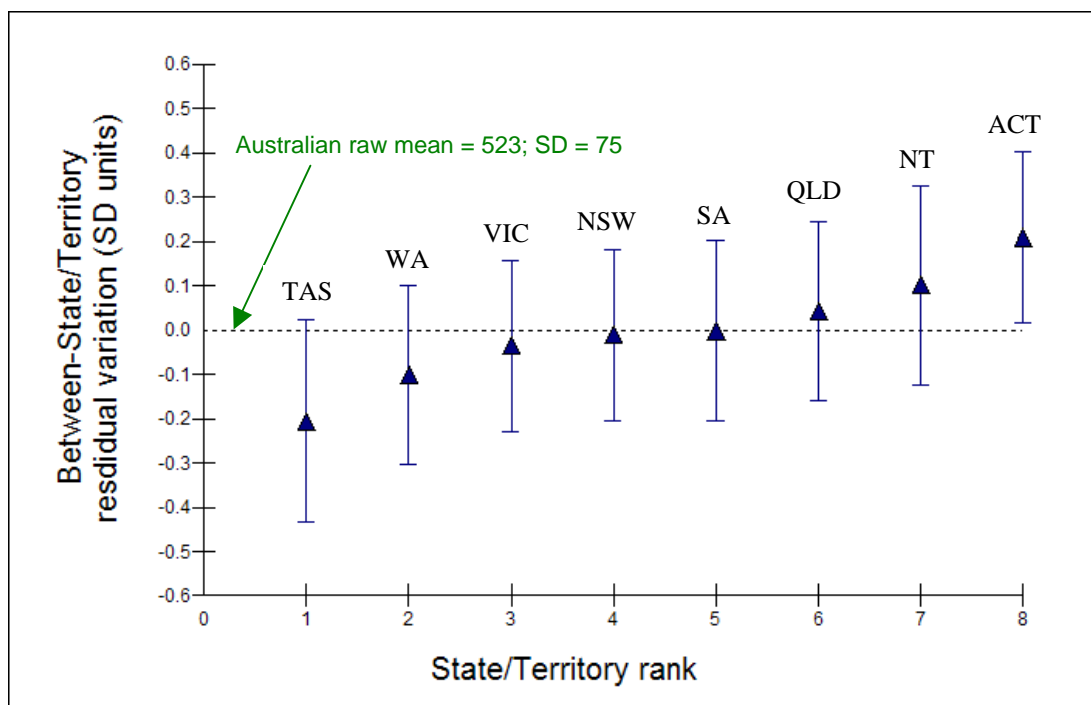


Figure A3.6 Ranked State/Territory raw residuals for TIMSS 2003 *Science* scores, showing mean-point estimates bounded by 95% 'uncertainty' intervals

Comment: With the exception of the ACT (above), the 'uncertainty' intervals around the unadjusted means for other States all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level in students' TIMSS 2003 *Science* scores – accounting for a non-significant 2.3% of the residual variance.

In the following model, adjustments are made for the 'intake' variables of *Gender*, *Age*, family *SES* (at the student-level), *Home Educational Resources* (HEDRES), and school average *SES* at the school-level (i.e., *ScAvSES* – to estimate the within-school average effect of *SES*, over-and-above that operating at the individual student-level). The results of the fitted model to the normalized data are given below, indicating the magnitude of the parameter estimates for the fitted variables (in SD units), and their respective standard errors given in parentheses.

$$\text{Science}_{ijk} = \beta_{0ijk} \text{Cons} + -0.205(0.025) \text{Gender}_{ijk} + -0.069(0.015) \text{Age}_{ijk} + \\ 0.074(0.019) \text{SES}_{ijk} + 0.156(0.020) \text{HEDRES}_{ijk} + 0.207(0.036) \text{ScAvSES}_{jk} \\ \beta_{0ijk} = 0.058(0.070) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.026(0.019)] \quad \text{Between-State/Territory res. var: 3.0\% (n.s.)}$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.237(0.027)] \quad \text{Between-schools res. var: 27.7\%}$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [0.593(0.013)] \quad \text{Within-schools res. var: 69.3\%}$$

Comment: These results indicate significant effects for: *Gender* (in favour of males), *Age* (in favour of younger students), both *SES* at the student-level and *ScAvSES* at the school-level, as well as *Home Educational Resources* (HEDRES), were significant predictors of Australian students' TIMSS 2003 *Science* achievement scores.

Together, the five fitted variables accounted for 15.2% of the variance in students' *Science* achievement scores, with an insignificant 3.0% of the residual variance at the State/Territory-level, a significant 27.7% of the residual variance due to variation between schools, and 69.3% at the student-level. Mean-adjusted residual plots at the State/Territory-level illustrate these findings in Figure A3.7 below.

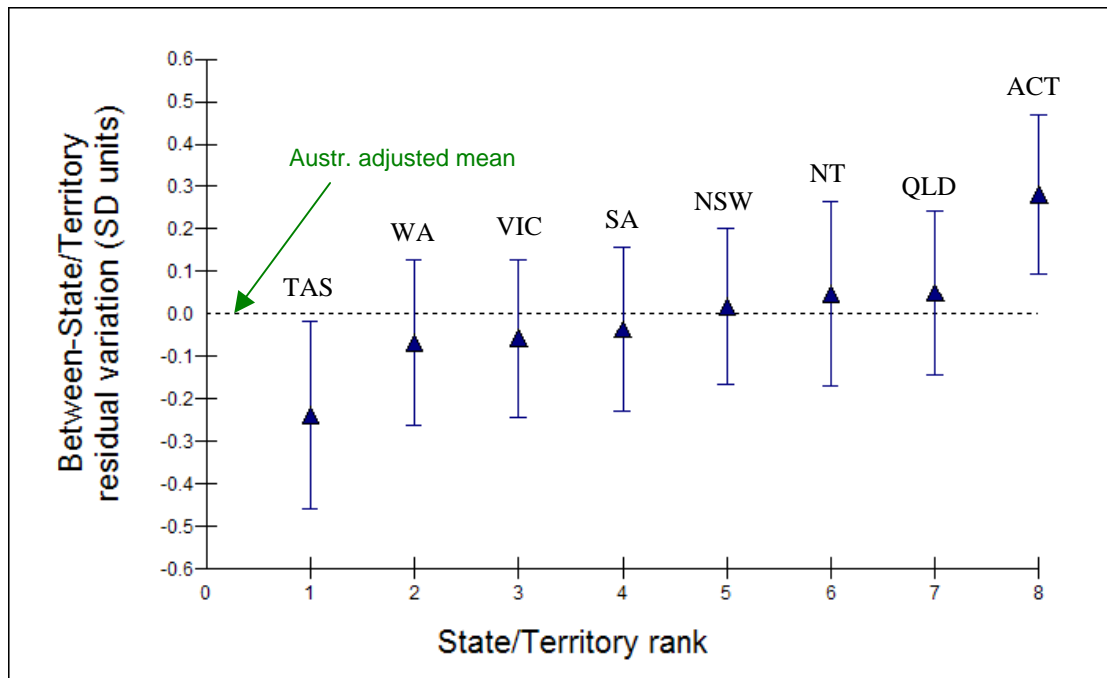


Figure A3.7 Plot of ranked State/Territory residuals, showing adjusted mean-point TIMSS 2003 *Science* score estimates, bounded by 95% 'uncertainty' intervals

Comment: From Figure A3.7, with the exception of TAS (marginally below) and ACT (above), the 95% 'uncertainty' intervals around the adjusted means for the other States and Territories all overlap the 'population' mean (zero) – indicating non-significant differences at the State/Territory-level (i.e., an insignificant 3% of the residual variance in students' TIMSS 2003 *Science* scores). While on average, students in ACT school performed significantly above the adjusted national mean and better than TAS students, there were no significant differences

between the adjusted mean performances of students located in WA, VIC, SA, NSW, NT, QLD and ACT schools, since their respective ‘uncertainty’ intervals overlap.

Although these results are of minor interest, they mask the variation between-schools at the State/Territory-level, as well as within each of the States and Territories separately. Following in Figure A3.8 is a plot of ranked residuals for 207 Australian schools from a multilevel analysis of residuals for performance in TIMSS 2003 *Science* showing adjusted mean-point estimates bounded by 95% ‘uncertainty’ intervals.

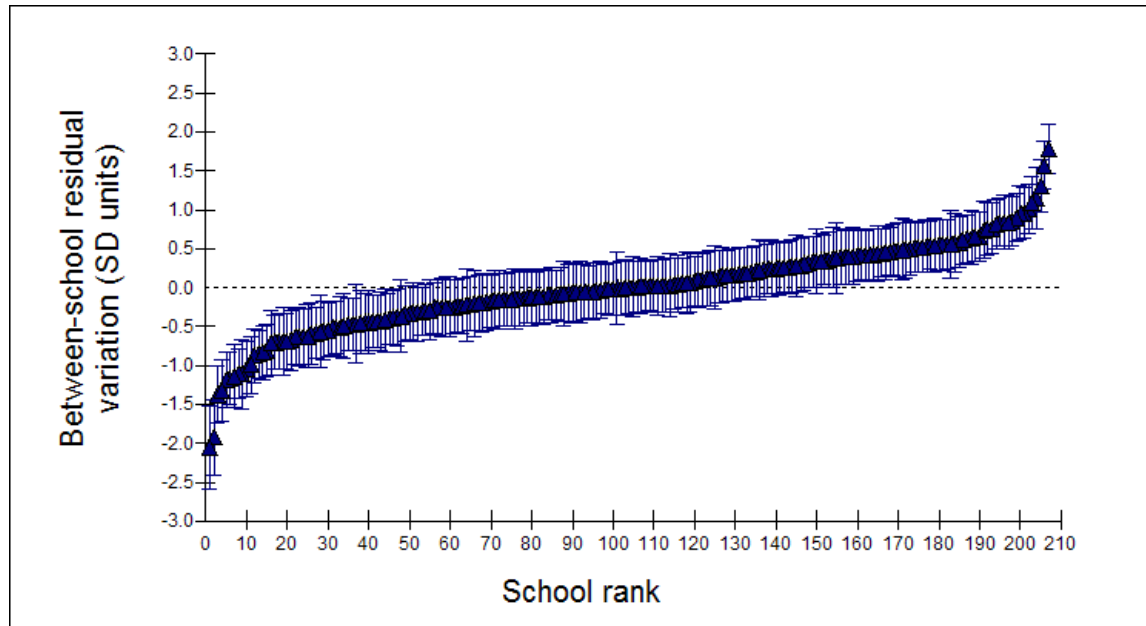


Figure A3.8 Plot of ranked residuals for 207 schools, showing adjusted mean-point TIMSS 2003 *Science* score estimates, bounded by 95% ‘uncertainty’ intervals

Comment: From Figure A3.8, the ‘uncertainty’ intervals around the adjusted means for approximately 50% the 207 schools all overlap the national mean (zero) – indicating non-significant between-school differences in students’ TIMSS 2003 Mathematics scores for these schools. The remaining ~50% of schools yielded student performances either significantly above or below the national mean, resulting in a large and significant between-school residual variation (27.7%). Again, these findings indicated differences in the quality of teaching and learning provision among Australian schools.